bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

FBXL19 recruits CDK-Mediator to the CpG islands of developmental genes to prime them for activation during lineage commitment

Emilia Dimitrova[1], Takashi Kondo[2*], Angelika Feldmann[1*], Manabu Nakayama[4], Yoko Koseki[2], Rebecca Konietzny[5,6], Benedikt M Kessler[5], Haruhiko Koseki[2,3], and Robert J Klose[1†]

1. Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom

2. Laboratory for Developmental Genetics, RIKEN Center for Integrative Medical Sciences, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

3. CREST, Japan Science and Technology Agency

4. Department of Technology Development, Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

5. TDI Mass Spectrometry Laboratory, Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

6. Present address: Agilent Technologies, Hewlett-Packard-Str. 8 - 76337 Waldbronn, Germany

*these authors contributed equally to this work

† Corresponding author: rob.klose@bioch.ox.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

CpG islands are gene regulatory elements associated with the majority of mammalian promoters, yet how they regulate gene expression remains poorly understood. Here, we identify FBXL19 as a CpG island-binding protein in mouse embryonic stem (ES) cells and show that it associates with the CDK-Mediator complex. We discover that FBXL19 recruits CDK-Mediator to CpG island-associated promoters of non-transcribed developmental genes to prime these genes for activation during cell lineage commitment. We further show that recognition of CpG islands by FBXL19 is essential for mouse development. Together this reveals a new CpG island-centric mechanism for CDK-Mediator recruitment to developmental gene promoters in ES cells and a requirement for CDK-Mediator in priming these developmental genes for activation during cell lineage commitment.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Multicellular development relies on the formation of cell type-specific gene expression programmes that support differentiation. At the most basic level, these expression programmes are defined by cell signalling pathways that control how transcription factors bind DNA sequences in gene regulatory elements and shape RNA polymerase II (RNAPolII)-based transcription from the core gene promoter (reviewed in (Spitz and Furlong, 2012)). In eukaryotes, the activity of RNAPolII is also regulated by a large multi-subunit complex, Mediator, which can directly interact with transcription factors at gene regulatory elements and with RNAPolII at the gene promoter to modulate transcription activation. The Mediator complex functions through regulating pre-initiation complex formation and controlling how RNAPolII initiates, pauses, and elongates. Therefore, Mediator is central to achieving appropriate transcription from gene promoters (reviewed in (Allen and Taatjes, 2015; Malik and Roeder, 2010)).

In mammalian genomes, CpG dinucleotides are pervasively methylated and this epigenetically maintained DNA modification is generally associated with transcriptional inhibition, playing a central role in silencing of repetitive and parasitic DNA elements (Klose and Bird, 2006; Schubeler, 2015). Most gene promoters, however, are embedded in short elements with elevated CpG dinucleotide content, called CpG islands, which remain free of DNA methylation (Bird et al., 1985; Illingworth and Bird, 2009; Larsen et al., 1992). Interestingly, mammalian cells have evolved a DNA binding domain, called a ZF-CxxC domain, which can recognize CpG dinucleotides when they are non-methylated (Lee et al., 2001; Voo et al., 2000). This endows ZF-CxxC domain-containing proteins with the capacity to recognize and bind CpG islands throughout the genome (Blackledge et al., 2010; Thomson et al., 2010). There are twelve mammalian proteins encoding a ZF-CxxC domain, most of which are found in large chromatin modifying complexes that post-translationally modify histone proteins to regulate

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

gene expression from CpG islands (Long et al., 2013a). This has led to the proposal that CpG islands may function through chromatin modification to affect transcription (Blackledge et al., 2013).

One of the first characterised ZF-CxxC domain-containing proteins was lysine specific demethylase 2A (KDM2A) (Blackledge et al., 2010). KDM2A is a JmjC-domain-containing histone lysine demethylase that removes histone H3 lysine 36 mono- and di- methylation (H3K36me1/2) from CpG islands (Blackledge et al., 2010; Tsukada et al., 2006). Like DNA methylation, H3K36me1/me2 is found throughout the genome and is thought to be repressive to gene transcription (Carrozza et al., 2005; Keogh et al., 2005; Peters et al., 2003). Therefore, it was proposed that KDM2A counteracts H3K36me2-dependent transcriptional inhibition at CpG island-associated gene promoters (Blackledge et al., 2010). Sequence-based homology searches revealed that KDM2A has two paralogues in vertebrates, KDM2B and FBXL19 (Katoh and Katoh, 2004). KDM2B, like KDM2A, binds to CpG islands and can function as a histone demethylase for H3K36me1/2 via its JmjC domain (Farcas et al., 2012; He et al., 2008). However, unlike KDM2A, it physically associates with the polycomb repressive complex 1 (PRC1) to control how transcriptionally repressive polycomb chromatin domains form at a subset of CpG island-associated genes (Blackledge et al., 2014; Farcas et al., 2012; He et al., 2013; Wu et al., 2013). These observations have suggested that, despite extensive similarity between KDM2A and KDM2B and some functional redundancy in histone demethylation, individual KDM2 paralogues have evolved unique functions. FBXL19 remains the least well characterized and most divergent of the KDM2 paralogues. Unlike KDM2A and KDM2B, it lacks a JmjC domain and, therefore, does not have demethylase activity (Long et al., 2013a). Previous work on FBXL19 has suggested that it plays a role in a variety of cytoplasmic processes that affect cell proliferation, migration, apoptosis, TGFβ signalling, and regulation of the innate immune response (Dong et al., 2014; Wei et al., 2013; Zhao et al., 2013; Zhao et

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

al., 2012). More recently, it has also been proposed to have a nuclear function as a CpG island-binding protein (Lee et al., 2017), but its role in the nucleus still remains poorly defined.

Here we investigate the function of FBXL19 in mouse embryonic stem (ES) cells and show that it is predominantly found in the nucleus where it localizes in a ZF-CxxC-dependent manner to CpG island promoters. Biochemical purification of FBXL19 revealed an association with the CDK-containing Mediator complex and we discover that FBXL19 can target CDK-Mediator to chromatin. Conditional removal of the CpG island-binding domain of FBXL19 in ES cells leads to a reduction in the occupancy of CDK8 at CpG islands associated with inactive developmental genes. FBXL19 and CDK-Mediator appear to play an important role in priming genes for future expression, as these genes are not appropriately activated during ES cell differentiation when the CpG island-binding capacity of FBXL19 is abolished or CDK-Mediator is disrupted. Consistent with an important role for FBXL19 in supporting normal developmental gene expression, removal of the ZF-CxxC domain of FBXL19 leads to perturbed development and embryonic lethality in mice. Together, our findings uncover an interesting new mechanism by which CDK-Mediator is recruited to gene promoters in ES cells and demonstrate a requirement for FBXL19 and CDK-Mediator in priming developmental gene expression during cell lineage commitment.

Results

FBXL19 is enriched in the nucleus and binds CpG islands via its ZF-CxxC domain.

FBXL19 shares extensive sequence similarity and domain architecture with the other KDM2 paralogues which are predominantly nuclear proteins (Figure 1A, (Blackledge et al., 2010; Farcas et al., 2012)), yet the function of FBXL19 has largely been described in the context of cytoplasmic processes (Dong et al., 2014; Wei et al., 2013; Zhao et al., 2013; Zhao et al., 2012).

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Therefore, we first examined the subcellular distribution of FBXL19 to understand whether it was unique amongst KDM2 paralogues in localizing to the cytoplasm. To achieve this, we generated a mouse ES cell line expressing epitope-tagged FBXL19 (FS2-FBXL19) and carried out immuno-fluorescent staining. This revealed that FBXL19 was almost exclusively nuclear (Figure 1B). When we carried out subcellular biochemical fractionation, FBXL19 was also enriched in the nuclear fractions in agreement with the immuno-fluorescent staining (Figure 1C). Importantly, FBXL19 was found not only in the soluble nuclear extract but also in the insoluble nuclear pellet, which contains chromatin-bound factors (Figure 1C). This suggested that FBXL19 may associate with chromatin, like other ZF-CxxC domain-containing proteins.

Based on the nuclear localization of FBXL19 and the fact that it encodes a highly conserved ZF-CxxC domain (Figure S1A), we set out to determine whether FBXL19 is a CpG island-binding protein. To achieve this, we carried out chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) for epitope-tagged FBXL19 and compared its binding profile with those we have previously generated for KDM2A and KDM2B in mouse ES cells (Blackledge et al., 2014; Farcas et al., 2012). A visual examination of FBXL19 ChIP-seq signal revealed that it was highly enriched at both computationally predicted CpG islands and genomic regions that contain BioCAP signal, an experimental measure of non-methylated DNA (Blackledge et al., 2012) (Figure 1D). We then identified all non-methylated CpG islands (NMIs) using BioCAP data (n=27698) (Long et al., 2013b) and extended our analysis across the genome. We observed that FBXL19 ChIP-seq signal at NMIs was highly similar to that of KDM2A and KDM2B (Figure 1E). Furthermore, FBXL19 ChIP-seq signal correlated well with the density of non-methylated CpG dinucleotides in NMIs, similarly to KDM2A and KDM2B, (Figure 1F, Figure S1B, and Figure S1C), and almost all FBXL19-occupied sites fell within NMIs (Figure S1D). Together these findings demonstrate that FBXL19 is a nuclear

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

143 protein that binds to CpG islands, an observation supported by a recent independent study (Lee
144 et al., 2017).

145 KDM2A and KDM2B rely on defined residues in their ZF-CxxC domain to recognize non-
146 methylated cytosine and bind CpG islands (Blackledge et al., 2010; Farcas et al., 2012). To
147 determine whether the association of FBXL19 with CpG islands is dependent on this domain,
148 we generated ES cell lines expressing either a mutant version of FBXL19, in which a key lysine
149 residue was substituted to alanine (K49A, Figure S1A) to disrupt the recognition of non-
150 methylated CpGs, or a truncated version of FBXL19, where the ZF-CxxC domain was deleted
151 (ΔCXXC) (Figure 1G). Importantly, the expression levels of wild type (WT) and mutant
152 FBXL19 transgenes were highly similar (Figure S1E). We then carried out ChIP-seq for
153 epitope-tagged FBXL19 in these cell lines and compared the binding profiles of the mutant
154 FBXL19 to that of WT FBXL19 (Figure 1H). This revealed a near complete loss of FBXL19
155 binding to chromatin when the ZF-CxxC domain was deleted and a slightly less dramatic effect
156 in the K49A mutant (Figure 1H, 1I, S1F, and S1G). Together, these observations demonstrate
157 that binding of FBXL19 to CpG islands relies on an intact and functional ZF-CxxC domain.

158

159 **FBXL19 interacts with the CDK-Mediator complex in ES cells.**

160 Our ChIP-seq analyses demonstrated that FBXL19 is targeted to CpG islands in a manner that
161 is highly similar to KDM2A and KDM2B (Figure 1). Although KDM2A and KDM2B localise
162 to the same genomic regions and show a high degree of sequence conservation, they associate
163 with different proteins (Farcas et al., 2012). This raised the interesting possibility that FBXL19
164 might also have unique interaction partners. To investigate this, we affinity-purified epitope-
165 tagged FBXL19 from ES cell nuclear extract and identified associated proteins by mass
166 spectrometry (AP-MS, Figure 2A and S2A). This revealed that FBXL19 interacts with SKP1,

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

a known F-box-binding protein that also associates with KDM2A and KDM2B (Bai et al., 1996; Farcas et al., 2012), and the nuclear proteasome activator PSME3 (Wojcik et al., 1998). Interestingly, we also identified multiple subunits of the Mediator complex that appeared to interact with FBXL19 in a sub-stoichiometric manner (Figure 2A and B). Biochemical purifications of Mediator have identified two distinct assemblies (Liu et al., 2001; Mittler et al., 2001; Taatjes et al., 2002). The first is characterized by the presence of the MED26 subunit which associates with the middle region of Mediator and this form of the complex interacts with RNAPolII (Naar et al., 2002; Paoletti et al., 2006; Sato et al., 2004; Takahashi et al., 2011). Alternatively, a kinase module, composed of CDK8/CDK19, MED12/12L, MED13/13L, and CCNC, can bind to Mediator in a manner which is mutually exclusive with MED26 (Taatjes et al., 2002). Interestingly, our FBXL19 purification exclusively yielded an interaction with the kinase-containing Mediator complex (CDK-Mediator), with no evidence for MED26 or RNAPolII inclusion (Figure 2). CDK8 and its paralogue CDK19 share 77% amino acid identity (89% similarity) (Audetat et al., 2017) and four out of the five peptides identified by mass spectrometry were common between the two proteins (data not shown). Therefore, it is likely that FBXL19 is able to interact with both CDK8- and CDK19-containing Mediator complexes. Importantly, interaction with CDK-Mediator was also evident when we performed affinity-purification of endogenous FBXL19 (Figure S2B-D). However, reciprocal immunopurifications of MED12 and CDK8 failed to yield detectable FBXL19 by Western blot (Figure S2D). This is in agreement with our mass spectrometry analysis, which indicated that the association of FBXL19 with CDK-Mediator is sub-stoichiometric, and suggests that this interaction is likely weak, as opposed to stable, *in vivo*.

We next wanted to determine which region of FBXL19 is required for interaction with CDK-Mediator. To do so, we transiently expressed full length FBXL19 or versions of FBXL19 with individual domains removed and performed affinity purification followed by Western blot

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

192 analysis. Intact FBXL19 and a version with the ZF-CxxC domain removed interacted with
193 CDK-Mediator, whereas removing the F-box domain resulted in a loss of this interaction
194 (Figure S2E). Therefore, FBXL19 relies on its F-box, and not its capacity to bind non-
195 methylated DNA, for its association with CDK-Mediator.

196 Based on a candidate approach, it was recently reported that FBXL19 could interact the
197 RNF20/40 E3 ubiquitin ligase in ES cells and regulate histone H2B lysine 120 ubiquitylation
198 (H2BK120ub1) (Lee et al., 2017). In our unbiased biochemical purification of FBXL19, we
199 did not identify an interaction with RNF20/40 by AP-MS or by Western blot analysis (Figure
200 2A and S2F). Furthermore, we failed to observe any relationship between the ability of
201 FBXL19 to associate with CpG islands and the levels of H2BK120ub1 (Figure S2G).
202 Therefore, the relevance of this proposed interaction remains unclear.

203 FBXL19 contains conserved F-box and leucine-rich repeat (LRR) domains. F-box proteins are
204 known to function as scaffolds and substrate recognition modules for the SKP1-Cullin-F-box
205 (SCF) protein complexes that ubiquitylate proteins for degradation by the proteasome (Skaar
206 et al., 2014; Skowyra et al., 1997). Based on the association of FBXL19 with SKP1 and PSME3
207 we speculated that FBXL19 might function as a SCF substrate-selector for CDK-Mediator, as
208 has previously been observed for the related F-box-containing protein FBW7 (Davis et al.,
209 2013). Interestingly, however, in our FBXL19 purifications, we did not detect the SCF complex
210 components CUL1 and RBX1/2, which are required for ubiquitin E3 ligase activity (Skaar et
211 al., 2014) (Figure 2A). Nevertheless, we investigated in more detail whether FBXL19 might
212 regulate CDK-Mediator protein levels. Treatment of ES cells with the proteasome inhibitor
213 MG132, which sequesters SCF substrates on their substrate selector, did not lead to elevated
214 levels of Mediator subunits in FBXL19 purifications as identified by AP-MS (unpublished
215 observation). In addition, transient overexpression of FBXL19 did not cause an appreciable
216 reduction in CDK-Mediator protein (Figure S2H). Based on these observations we conclude

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that FBXL19 does not function as a SCF substrate selector for CDK-Mediator. This raised the interesting possibility that FBXL19 may function in a proteasome-independent manner with CDK-Mediator at CpG islands.

FBXL19 recruits CDK-Mediator to chromatin.

It has previously been shown that transcription factors can recruit the Mediator complex to enhancers and gene promoters (Poss et al., 2013), yet the complement of mechanisms by which Mediator is targeted to chromatin remains very poorly defined. Given that FBXL19 does not appear to regulate CDK-Mediator protein levels, we hypothesized that it might instead function to recruit CDK-Mediator to chromatin. To test this possibility, we took advantage of a synthetic system we have developed to nucleate proteins *de novo* on chromatin and test their capacity to recruit additional factors (Blackledge et al., 2014). Fusion of the Tet repressor DNA-binding domain (TetR) to FBXL19 allowed the recruitment of FBXL19 to a short array of Tet repressor DNA binding sites (TetO), engineered into a single site on mouse chromosome 8 ((Blackledge et al., 2014), Figure 3A). To determine whether FBXL19 was sufficient to recruit CDK-Mediator to chromatin, we stably expressed TetR or the TetR-FBXL19 fusion protein in the TetO array-containing ES cell line (Figure S3A). We then carried out ChIP for CDK-Mediator subunits (Figure 3B). Consistent with our biochemical observations (Figure 2), when we examined the binding of CDK8 and MED12 over the TetO array, we observed an enrichment in the TetR-FBXL19 line. Furthermore binding of MED4, which is part of the core Mediator complex, was also evident at the TetO array (Figure 3B). This suggests that FBXL19 may be sufficient to recruit a holo-CDK-Mediator complex, in keeping with its biochemical co-purification with multiple components of both the CDK module and core Mediator complex (Figure 2). Importantly, stable expression of TetR-KDM2A and TetR-KDM2B did not lead to CDK8 recruitment, indicating that this activity is unique to FBXL19 (Figure S3B). Further

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

work will be required to determine the dynamics with which individual Mediator components are recruited to chromatin by FBXL19 as well as the precise composition of such complexes. Together these observations demonstrate that FBXL19 interacts specifically with the CDK-Mediator and can recruit this complex *de novo* to chromatin.

FBXL19 is required for appropriate CDK8 occupancy at a subset of CpG island-associated promoters.

FBXL19 binds specifically to CpG islands (Figure 1) and can recruit CDK-Mediator to an artificial binding site on chromatin (Figure 3). This raised the interesting possibility that FBXL19 may help to recruit CDK-Mediator to CpG islands in ES cells. To examine this, we first performed CDK8 ChIP-seq in ES cells and compared it to FBXL19 ChIP-seq. Unlike FBXL19, CDK8 occupancy was not restricted to CpG islands (Figure S4A), with only 67.5% of CDK8 peaks overlapping with NMIs and CDK8 binding showing a limited correlation with BioCAP signal (Spearman correlation – 0.48, Figure S4B). This is in line with previous studies demonstrating that the Mediator complex is recruited to both enhancers and gene promoters (Kagey et al., 2010; Malik and Roeder, 2010). Interestingly, however, we observed enrichment of CDK8 at FBXL19 peaks (Figure S4C) raising the possibility that FBXL19 may contribute to its occupancy at these sites. To directly test this, we developed an ES cell system in which the exon encoding the ZF-CxxC domain of FBXL19 is flanked by loxP sites (*Fbxl19*$^{fl/fl}$) (Figure S4D) and which expresses tamoxifen-inducible ERT2-Cre recombinase. Upon addition of tamoxifen (OHT), the ZF-CxxC-encoding exon is excised, yielding a form of FBXL19 that lacks the ZF-CxxC domain (FBXL19$^{\Delta CXXC}$, Figure 4A, 4B, and S4E) and can, therefore, no longer bind CpG islands (Figure 1). This model cell system allows us to specifically examine the CpG island-associated functions of FBXL19 without affecting its other proposed roles (Dong et al., 2014; Wei et al., 2013; Zhao et al., 2013; Zhao et al., 2012). Following removal

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of the ZF-CxxC domain of FBXL19, we observed some reductions in FBXL19 protein levels (Figure 4B), but importantly CDK8 levels were unaffected (Figure S4F). Using this conditional *Fbxl19*$^{\Delta CXXC}$ ES cell line, we then examined CDK8 occupancy on chromatin by ChIP-seq before and after OHT treatment. Genome-wide profiling of CDK8 in *Fbxll19*$^{\Delta CXXC}$ ES cells did not reveal widespread alterations in CDK8 binding (Figure 4C left and S4G), indicating that FBXL19 is not the central determinant driving CDK8 recruitment to most of its binding sites in the genome. Intriguingly, however, a more detailed site-specific analysis revealed a subset of CDK8 target sites that displayed significant alteration in CDK8 occupancy (Figure 4C, 4D, and S4G). In keeping with a potential role for FBXL19 in CDK8 recruitment, the majority of the affected sites showed reduced CDK8 binding (n=783) (Figure 4C).

Upon closer examination of sites with reduced CDK8 occupancy in *Fbxll19*$^{\Delta CXXC}$ ES cells, we discovered that they tended to coincide with broad regions of CDK8 enrichment (Figure 4E and F), a feature often associated with super-enhancers (Whyte et al., 2013). However, a comparison of these sites to the location of super-enhancers in mouse ES cells indicated that these were distinct (Figure 4G). Instead, sites displaying reduction in CDK8 occupancy coincided with broad regions of non-methylated DNA (Figure 4D, 4H, and S4G), tended to be associated with gene promoters (Figure S4H), and had elevated levels of FBXL19 (Figure 4I, S4G, S4H, and S4I). Together, these observations suggest that although binding of CDK8 to most of its target sites in the genome is achieved independently of FBXL19, a subset of broad CpG island-associated gene promoters appear to rely on FBXL19 for appropriate CDK8 occupancy.

FBXL19 targets CDK8 to promoters of silent developmental genes in ES cells.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Despite associating widely with CpG island gene promoters, FBXL19 appears to play a very specific role in maintaining appropriate CDK8 occupancy at a particular subset of broad CpG island-associated promoters (Figure 4). We were therefore curious to know whether something distinguishes these gene promoters from other FBXL19-bound promoters, where FBXL19 does not contribute appreciably to CDK8 occupancy (Figure S4G). To achieve this, we initially performed gene ontology (GO) analysis (Huang da et al., 2009) on the genes with reduced CDK8 levels at their promoters in *Fbxl19*$^{\Delta CXXC}$ ES cells. Interestingly, this revealed that these genes were strongly enriched for developmental processes (Figure 5A) in agreement with our previous discovery that broad CpG islands are an evolutionary conserved feature of developmentally regulated genes (Long et al., 2013b). In contrast, genes associated with an increased binding of CDK8 did not show any significant GO term enrichment (unpublished observation).

In pluripotent mouse ES cells, many developmental genes are inactive and only become expressed as cells commit to more differentiated lineages (Boyer et al., 2006). Importantly, the genes that exhibited reductions in CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ cells were expressed at significantly lower levels than most other genes in ES cells (Figure 5B). Previous work has identified a subset of poised but inactive developmental genes in ES cells that are proposed to exist in a bivalent chromatin state characterised by the co-occurrence of histone H3 lysine 4 and 27 methylation (H3K4/K27me) (Bernstein et al., 2006; Mikkelsen et al., 2007). Therefore, we examined whether sites that rely on FBXL19 for normal CDK8 binding also corresponded to bivalent regions in ES cells. We found that these regions are enriched for H3K27me3 (Figure S5A) and have elevated H3K4me3 compared to other H3K27me3 modified sites that lack CDK8 (Figure S5B). Therefore, sites that rely on FBXL19 for normal CDK8 binding and correspond to silent developmental gene promoters are also bivalent. To ask whether these genes are induced during cell lineage commitment, we compared their expression levels in ES

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

315 cells and following retinoic acid (RA) treatment which induces differentiation (Figure S6B and
316 S6C). This clearly demonstrated that the genes which rely on FBXL19 for appropriate CDK8
317 binding in the ES cell state can become transcriptionally activated during stem cell lineage
318 commitment (Figure 5C).

319 The observation that FBXL19 appears to be important for CDK8 occupancy at largely inactive
320 genes was intriguing given that previous work characterising Mediator function has usually
321 focussed on its activity at actively transcribed or induced genes (reviewed in (Poss et al.,
322 2013)). We therefore set out to examine the relationship between FBXL19, CDK8 and gene
323 expression in more detail. We first separated all promoters based on expression of their
324 associated gene (Figure 5D and S5D). We observed that CDK8 occupancy was in general
325 linked to gene activity (Figure 5E), in agreement with previous reports (Alarcon et al., 2009;
326 Bancerek et al., 2013; Donner et al., 2010; Donner et al., 2007; Fryer et al., 2004; Galbraith et
327 al., 2013). However, at FBXL19-bound gene promoters, CDK8 occupancy was similar in both
328 the lowly and highly expressed subsets of genes (Figure 5E). Importantly, CDK8 occupancy at
329 promoters of lowly expressed genes was reduced in *Fbxl19*$^{\Delta CXXC}$ ES cells (Figure 5F).
330 Therefore, these observations reveal that CDK8 is enriched at promoters of inactive or lowly
331 expressed genes in ES cells and its binding is dependent on recognition of CpG islands by
332 FBXL19.

333

334 **Removing the CpG island-binding domain of FBXL19 results in a failure to induce**
335 **developmental genes during ES cell differentiation.**

336 FBXL19 appears to play a role in recruiting CDK8 to a class of genes that are repressed in the
337 ES cell state and become activated during cell linage commitment (Figure 5). However, it
338 remained unclear whether FBXL19 is required for the activation of these developmental genes.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To address this question, we induced differentiation of *Fbxl19*$^{fl/fl}$ and *Fbxl19*$^{\Delta CXXC}$ ES cells with RA (Figure 6A) and compared the expression of several genes which showed reduced CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells (Figure 6B). We observed no significant differences in the expression of these genes in wild type *Fbxl19*$^{fl/fl}$ and *Fbxl19*$^{\Delta CXXC}$ ES cells where these genes are silent. Strikingly however, following RA treatment, *Fbxl19*$^{\Delta CXXC}$ cells failed to induce these genes appropriately (Figure 6B). Similarly, developmental genes were not appropriately induced during embryoid body differentiation of *Fbxl19*$^{\Delta CXXC}$ ES cells (Figure S6A). Importantly, this demonstrates that FBXL19 is required for the appropriate activation of developmental gene expression during cell lineage commitment.

To understand the extent to which genes are not appropriately induced during differentiation of *Fbxl19*$^{\Delta CXXC}$ ES cells, we examined ongoing transcription in both the ES cell state and after RA-mediated differentiation using short time-scale 4-thiouridine-based labelling and RNA sequencing (4sU RNA-seq) (Figure 6A). We observed a significant number of genes that become induced upon RA treatment in wild type cells (n=4051, Figure S6B). GO term analysis confirmed that these genes are associated with ES cell differentiation and early embryonic development (Figure S6C). We then used differential gene expression analysis to compare transcription in *Fbxl19*$^{fl/fl}$ and *Fbxl19*$^{\Delta CXXC}$ cells (Figure 6C). While very few significant changes in gene expression were observed in the ES cell state, following RA-induced differentiation we identified a large number of genes (n=552) that had significantly lower expression levels in *Fbxl19*$^{\Delta CXXC}$ cells (Figure 6C). This is consistent with our results when examining individual genes (Figure 6B and S6A) and with a recent study where the expression of a selection of genes was examined following FBXL19 knock-down (Lee et al., 2017). GO analysis revealed that the set of genes that were not appropriately activated were associated with genes involved in developmental processes (Figure 6D), unlike genes with increased expression which were not associated with these processes (Figure S6D). Consistent with our

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

364 analysis of FBXL19-dependent CDK8 occupancy at promoters of developmental genes (Figure
365 5B and 5C), these genes were lowly expressed in wild type ES cells (Figure 6E) and tended to
366 become activated following RA induction (Figure 6F). Importantly, these genes exhibited a
367 reduction in CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells (Figure 6G and S6E), supporting the idea
368 that FBXL19 recruits CDK8 to this subset of CpG island-associated developmental genes and
369 that this may prime these genes for activation during differentiation.

370

371 **FBXL19 target genes rely on CDK-Mediator for activation during differentiation**

372 Given that FBXL19 is required for CDK8 occupancy at the promoters of a series of silent
373 developmental genes and for their activation during differentiation (Figure 5 and 6), we
374 hypothesized that FBXL19 may prime these genes for future activation through the activity of
375 CDK-Mediator. To address this interesting possibility, we developed a system to inducibly
376 remove MED13 and its closely related paralogue MED13L in ES cells (*Med13/13L*$^{fl/fl}$ ERT2-
377 Cre ES cells, Figure 7A and S7A). We chose to inactivate MED13/13L as it has previously
378 been shown to physically link the CDK-kinase module to the core Mediator complex and
379 underpin the formation of a functional CDK-Mediator (Knuesel et al., 2009; Tsai et al., 2013).
380 Treatment of the *Med13/13L*$^{fl/fl}$ ES cells with OHT resulted in a loss of MED13 and MED13L
381 protein (MED13/13L KO) and reduced levels of the other subunits of the CDK-kinase module
382 (Figure 7B). Importantly, removal of MED13/13L also led to a loss of CDK8 binding to
383 chromatin (Figure 7C) but did not have an appreciable effect on the expression of FBXL19
384 target genes in the ES cell state (Figure 7D). We next induced differentiation of the
385 MED13/13L KO cells with RA (Figure 7A). Importantly, when we then analysed the
386 expression of a series of genes that rely on FBXL19 for activation during differentiation (Figure
387 7D), we observed that these genes also failed to appropriately induce in MED13/13L KO cells
388 (Figure 7D). Together this suggests that FBXL19, via its association with CpG islands,

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

389 recognises a subset of silent developmental genes in ES cells in order to recruit the CDK-
390 Mediator complex and prime these genes for future activation.

391

392 **Ablating the capacity of FBXL19 to bind CpG islands causes embryonic lethality.**

393 Our results suggest that FBXL19 contributes to gene activation during cell lineage commitment
394 via recognition of CpG islands and recruitment of the CDK-Mediator complex. Given that
395 *Fbxl19*$^{\Delta CXXC}$ cells display impaired gene activation in our *in vitro* differentiation model, we
396 asked whether the inability of FBXL19 to bind and function at CpG islands could also affect
397 mouse development. In order to address this, we generated *Fbxl19*$^{\Delta CXXC}$ mutant mice by
398 crossing *Fbxl19*$^{fl/fl}$ animals with animals constitutively expressing Cre recombinase. From
399 these crosses, we failed to obtain any viable *Fbxl19*$^{\Delta CXXC}$ homozygous offspring, indicating that
400 removal of the ZF-CxxC domain of FBXL19 leads to embryonic lethality (Figure 8A). We then
401 investigated at which stage *Fbxl19*$^{\Delta CXXC}$ homozygous embryos were affected and found
402 homozygous embryos were observed at normal Mendelian ratios until 10.5 days postcoitum
403 (dpc) (Figure 8A). At 9.5 dpc, gross embryonic morphology appeared to be intact but at 10.5
404 dpc (Figure 8B), the embryos exhibited a clear growth retardation and showed, to varying
405 extents, reduced elongation of the trunk, hypomorphic limb buds and cephalic region. This
406 included undeveloped facial mesenchyme, including defects in maxillary and mandibular
407 components of first branchial arches (indicated by yellow and red arrows in Figure 8B,
408 respectively), and hypomorphic cardiac mesenchyme. Some living *Fbxl19*$^{\Delta CXXC}$ homozygous
409 embryos, characterized by a beating heart, were observed at 12.5 dpc but they were of similar
410 size and external appearance to 10.5 dpc mutant embryos suggesting that normal development
411 had ceased by this point. Together our findings demonstrate that FBXL19, and its ability to
412 recognize CpG islands, is essential for normal mouse embryonic development.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Here we discover that FBXL19 recognizes CpG islands throughout the genome in a ZF-CxxC-dependent manner (Figure 1). Unlike other ZF-CxxC proteins which associate with chromatin-modifying complexes, we show that FBXL19 interacts with the CDK-Mediator complex (Figure 2). This uncovers an unexpected link between CpG islands and a complex that regulates gene expression through interfacing with the transcriptional machinery. We demonstrate that FBXL19 can recruit CDK-Mediator to chromatin (Figure 3) and, via recognition of CpG islands, plays an interesting role in supporting CDK8 occupancy at a subset of promoters associated with developmental genes, which are inactive in ES cells (Figure 4 and 5). At these CpG islands, FBXL19 and CDK-Mediator function to prime the associated genes for activation during ES cell differentiation (Figure 6 and 7). Consistent with an important role of FBXL19 in supporting normal developmental gene expression, removal of the ZF-CxxC domain of FBXL19 leads to perturbed development and embryonic lethality in mice (Figure 8). Together these new discoveries reveal that CpG islands and FBXL19 can interface with CDK-Mediator to orchestrate normal gene expression during lineage commitment.

Previously, another F-box protein, FBW7, was shown to function as an E3 ubiquitin ligase substrate selector for the MED13/13L subunit of CDK-Mediator and to regulate its stability through proteasomal degradation (Davis et al 2013). By controlling CDK-Mediator abundance, one could envisage how this might shape Mediator function in gene expression. FBXL19 also encodes an F-box domain and associates with SKP1, a central component of SCF E3 ubiquitin ligase complexes. However, we do not find evidence that FBXL19 regulates CDK-Mediator stability via the proteasome (Figure S2H). Instead, we discover that FBXL19 can recruit CDK-Mediator to chromatin and, more specifically, to CpG islands via its ZF-CxxC domain to support gene activation. Although we currently do not know the defined subunits and surfaces in CDK-Mediator that FBXL19 interacts with, our biochemical experiments indicate that this

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

438 relies on an intact F-box domain in FBXL19 (Figure S2E). It is tempting to speculate that
439 FBXL19 could interact directly with MED13/13L given their recognition by the related F-box
440 protein, Fbw7 (Davis et al., 2013). Nevertheless, a key observation in our work is that FBXL19
441 appears to have functionally diverged from other F-box proteins during vertebrate evolution by
442 acquiring a DNA-binding domain that allows it to recruit proteins to chromatin, instead of
443 targeting them for ubiquitylation. This general feature is shared with the FBXL19 paralogue,
444 KDM2B, which associates with and recruits the PRC1 complex to CpG islands (Farcas et al.,
445 2012; He et al., 2013; Wu et al., 2013). In agreement with our observations, a large-scale
446 proteomics screen previously suggested that an interaction between FBXL19 and Mediator
447 may also exist in cancer cells (Tan et al., 2013). Given that CDK8 can function as an oncogene
448 (Adler et al., 2012; Firestein et al., 2008; Morris et al., 2008), it will be interesting to understand
449 whether FBXL19 plays a role in targeting CDK-Mediator to CpG islands in non-embryonic
450 tissues, and whether this activity may promote CDK8-driven tumorigenesis.

451 Individual Mediator subunits have been shown to interact with DNA-binding transcription
452 factors that recruit the Mediator complex to specific DNA sequences in regulatory elements
453 and support gene expression (Black et al., 2006; Blazek et al., 2005; Fondell et al., 1996; Malik
454 and Roeder, 2010). In part, Mediator is thought to achieve this by bridging enhancer elements
455 to the core gene promoter and RNAPolII (Allen and Taatjes, 2015; Jeronimo et al., 2016;
456 Petrenko et al., 2016). However, in mammals these proposed mechanisms are extrapolated
457 from studying only a subset of transcription factors and genes. Therefore, the defined
458 mechanisms that shape Mediator occupancy on chromatin remain very poorly understood. Here
459 we provide evidence for a completely new gene promoter-associated CDK-Mediator targeting
460 mechanism that relies on FBXL19 and CpG island recognition. Surprisingly, although
461 FBXL19 localises broadly to CpG islands throughout the genome (Figure 1), we only observed
462 a reliance on FBXL19 for CDK8 binding at a subset of sites (Figure 4). Similarly, KDM2B

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

463 binds broadly to CpG islands, yet has a specificity in shaping PRC1 recruitment and polycomb
464 repressive chromatin domain formation at a subset of developmental genes in mouse ES cells
465 (Farcas et al., 2012). We have previously suggested that this could result from ZF-CxxC
466 domain-containing proteins broadly sampling CpG islands, with their activity or affinity for
467 certain sites being shaped by local gene activity or chromatin environment (Klose et al., 2013).
468 In keeping with these general ideas, FBXL19 is required for appropriate CDK8 binding at
469 bivalent genes in ES cells (Figure S5A and S5B). In future work, it will be important to
470 understand whether this unique chromatin state and/or the absence of transcriptional activity at
471 these sites define the requirement for FBXL19 in CDK8 recruitment, or whether FBXL19
472 targeting occurs at most CpG islands sites but is masked through redundancy with transcription
473 factor and gene activity-dependent targeting modalities.

474 Historically, CDK-Mediator has been associated with gene repression (Akoulitchev et al.,
475 2000; Chi et al., 2001; Elmlund et al., 2006; Hengartner et al., 1998; Knuesel et al., 2009; Pavri
476 et al., 2005). Therefore, we were not surprised to observe that CDK8 occupied the promoters
477 of repressed developmental genes in ES cells. While abrogating the CpG island-binding
478 activity of FBXL19 or deletion of MED13/13L resulted in reduced CDK8 binding at these
479 sites, it did not lead to an appreciable effect on gene expression in the ES cell state (Figure 6,
480 S6C, and 7D). This suggests that CDK-Mediator is not required for the repression of these
481 genes in ES cells. Instead, these genes fail to properly activate during differentiation (Figure 6,
482 S6, and 7D). These observations are in line with several reports that support the idea that CDK-
483 Mediator is also involved in gene activation (Alarcon et al., 2009; Donner et al., 2007; Hirst et
484 al., 1999; Donner et al., 2010; Galbraith et al., 2013). Therefore, in the context of mouse ES
485 cells, we propose that CDK-Mediator may be required to prime silent developmental genes for
486 future gene activation through a mechanism that relies on the recognition of CpG island
487 promoters by FBXL19. This priming could contribute to gene activation during differentiation

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

488 through one of the several mechanisms by which CDK-Mediator has been proposed to affect
489 gene transcription, including regulating RNAPolII pre-initiation complex assembly,
490 polymerase pausing/elongation, or through mediating long range interactions with distal
491 regulatory elements (Allen and Taatjes, 2015; Belakavadi and Fondell, 2010; Donner et al.,
492 2010; Galbraith et al., 2013; Kagey et al., 2010). Clearly, understanding the defined mechanism
493 by which FBXL19-dependent CDK-Mediator recruitment supports normal gene activation
494 during differentiation remains an important question for future work.

495 Our finding that FBXL19 can target CDK8 to gene promoters and that this is required to prime
496 the expression of developmental genes during ES cell differentiation is conceptually
497 reminiscent of recent work in yeast and human systems which suggested that CDK8 is required
498 for appropriate re-induction of inducible genes following an initial activation stimulus (D'Urso
499 et al., 2016). In this study, CDK8 was found to associate with the promoters of inducible genes
500 following initial activation, even in the absence of appreciable ongoing gene transcription,
501 thereby acting as a form of transcriptional memory (D'Urso et al., 2016). Interestingly, this
502 requirement for CDK8 in normal gene activation appears shared in the context of our
503 developmental gene induction paradigm. However, we find that, in the case of developmental
504 genes, previous gene activation may not be required for the recruitment of CDK8. Instead, this
505 could be achieved by FBXL19 directly recognising and targeting CDK8 to CpG islands.
506 Therefore, we prefer to view this as priming as opposed to memory. Nevertheless, collectively
507 these observations point to an important role for CDK-Mediator binding to gene promoters
508 prior to activation and as a way of supporting subsequent gene induction. This provides new
509 evidence that the roles of CDK-Mediator in gene regulation are more complicated than simply
510 conveying activation signals to RNAPolII through recruitment by transcription factors and
511 suggests it may have evolved to play a unique role in regulating gene expression in stem cells
512 and during development.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

513 **Material and Methods**

514 **Cell culture**

515 Mouse ES cells were cultured on gelatine-coated dishes in DMEM (Thermo scientific)
516 supplemented with 15% fetal bovine serum (BioSera), L-Glutamine, beta-mercaptoethanol,
517 non-essential amino acids, penicillin/streptomycin (Thermo scientific) and 10ng/mL leukemia-
518 inhibitory factor. *Fbxl19*$^{fl/fl}$ ES cells were treated with 800nM 4-hydroxytamoxifen (Sigma) for
519 96 hours in order to delete the Zf-CxxC domain. For RA differentiation of ES cells, $4x10^4$
520 cells/cm^2 were allowed to attach to gelatinised dishes (~12 hours) and treated with 1µM retinoic
521 acid (Sigma-Aldrich) in EC-10 medium (DMEM supplemented with 10% fetal bovine serum,
522 L-Glutamine, beta-mercaptoethanol, non-essential amino acids and penicillin/streptomycin)
523 for 72 hours. For embryonic body differentiation, $2x10^6$ cells were plated on non-adhesive
524 10cm dishes in EC-10 medium and cultured for the indicated days. For generation of stable
525 cell lines, E14 ES cells were transfected using Lipofectamine 2000 (Thermo scientific)
526 following manufacturer's instructions. Stably transfected cells were selected for 10 days using
527 1µg/ml puromycin and individual clones were isolated and expanded in the presence of
528 puromycin to maintain the transgene expression. TOT2N E14 cells used for TetR targeting
529 experiments were previously described (Blackledge et al., 2014). 293T cells were cultured in
530 EC-10 media. Transient overexpression of FBXL19 was performed by transfecting 293T cells
531 using Lipofectamine 2000 (Thermo scientific) followed by selection with 400ng/µl G418 for
532 48 hours. Proteasome inhibition was performed for 4 hours using 10µM MG132 inhibitor
533 (Sigma-Aldrich).

534

535 **Generation of *Fbxl19*$^{\Delta CXXC}$ conditional knockout mouse**

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

536 The targeting construct was generated from C57BL6/J mouse genomic sequence spanning
537 mm9 chromosome 7: 134,888,487-134,895,358 containing the exons 1 to 6 of *Fbxl19* genomic
538 region. Recombination was carried out by Gateway system (Life Technologies). One of the
539 loxP sequences was inserted at mm9 chr7:134,891,537, and FRT flanked PGK-neo was
540 inserted at mm9 chr7:134,892,000 together with an additional loxP sequence. The targeting
541 construct was electroporated into M1 ES cells to obtain targeted insertion. Clones of targeted
542 ES cells were aggregated with 8-cell embryos to generate the targeted mouse line. The
543 *Fbxl19*$^{fl/fl}$ line was generated by removal of the PGK-neo marker gene by mating the targeted
544 mice with mice expressing FLP recombinase. These *Fbxl19*$^{fl/fl}$ mice were further mated with
545 mice harbouring the *ROSA26-CreErt2* locus to generate *Fbxl19*$^{fl/fl}$:*ROSA26-CreErt2*$^{+/-}$ mice,
546 from which the *Fbxl19*$^{fl/fl}$ ES cells used in this study were derived.

547

548 **Generation of *Med13/13L*$^{fl/fl}$ conditional ES cells line**

549 In order to generate a conditional *Med13/13L*$^{fl/fl}$ ES cell line, we inserted loxP sites downstream
550 of exons 7 and upstream of exons 8 of the *Med13* and *Med13L* genes. The targeting constructs
551 for the insertion of each loxP site were designed to have 150bp homology arms flanking the
552 loxP site and to carry a mutated PAM sequence to prevent retargeting by the Cas9 enzyme.
553 The targeting constructs were purchased from GeneArt Gene Synthesis (Thermofisher
554 scientific). The pSpCas9(BB)-2A-Puro(PX459)-V2.0 vector was obtained by Addgene
555 (#62988). sgRNAs were designed to specifically target the desired genomic region for each
556 loxP insertion (http://crispor.tefor.net/crispor.py) and were cloned into the Cas9 vector as
557 previously described (Ran et al., 2013). ES cells that express the ERT2-Cre recombinase from
558 the *ROSA26* locus (*ROSA26:ERT2-Cre*) were used. First, the downstream loxP sites for *Med13*
559 and *Med13L* were targeted. *ROSA26:ERT2-Cre* ES cells were transiently co-transfected with
560 1µg of each Cas9-sgRNA plasmid and 3.5µg of each targeting construct using Lipofectamine

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3000 (ThermoFischer Scientific). Successfully transfected ES cells were selected for 48 hours with 1µg/ml puromycin. Individual clones were screened by genotyping PCR to identify correctly targeted homozygous clones. A clone homozygous for both Med13 and Med13L LoxP1 sites was then used to target the upstream loxP sites using the same transfection protocol and screening strategy. Correct loxP targeting was verified by sequencing of the genomic region surround the loxP sites.

DNA constructs

For generation of FBXL19 expression constructs, the full length, ΔCxxC or ΔF-box cDNA of mouse *Fbxl19* (IMAGE ID 6401846, Source Bioscience) was PCR amplified and inserted into a pCAG-IRES-FS2 vector (Farcas et al., 2012) via ligation-independent cloning (LIC). Mutation of the ZF-CxxC domain of FBXL19 (K49A) was generated via site-directed mutagenesis using the Quikchange mutagenesis XL kit (Stratagene). To generate TetR-FBXL19 fusion expression plasmid, *Fbxl19* cDNA was cloned into pCAG-FS2-TetR (Blackledge et al., 2014) via LIC. For transient overexpression in 293T cells, full length FBXL19 cDNA was cloned into pcDNA3-2xFlag vector by conventional cloning. All plasmids were sequence-verified by sequencing.

Nuclear extract preparation and immunoprecipitation

Cells were harvested by scraping in PBS at 4°C, resuspended in 10x pellet volume (PV) of Buffer A (10mM Hepes pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, 0.5mM DTT, 0.5mM PMSF, cOmplete protease inhibitor cocktail (Roche)) and incubated for 10 min at 4°C with slight agitation. After centrifugation, the cell pellet was resuspended in 3x PV Buffer A containing 0.1% NP-40 and incubated for 10 min at 4°C with slight agitation. Nuclei were recovered by

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

centrifugation and the soluble nuclear fraction was extracted for 1 hour at 4°C with slight agitation using 1x PV Buffer C (10mM Hepes pH 7.9, 400mM NaCl, 1.5mM $MgCl_2$, 26% glycerol, 0.2mM EDTA, cOmplete protease inhibitor cocktail). Protein concentration was measured using Bradford assay.

For small-scale co-immunoprecipitation, 600µg of nuclear extract was diluted in BC150 buffer (50mM Hepes pH 7.9, 150mM KCl, 0.5mM EDTA, 0.5mM DTT, cOmplete protease inhibitor cocktail). Samples were incubated with the respective antibodies and 25U benzonase nuclease overnight at 4°C. Protein A agarose beads (RepliGen) were blocked for 1 hour 4°C in Buffer BC150 containing 1% fish skin gelatine (Sigma) and 0.2 mg/ml BSA (NEB). The blocked beads were added to the samples and incubated for 4 hours at 4°C. Washes were performed using BC150 containing 0.02% NP-40. The beads were resuspended in 2x SDS loading buffer and boiled for 5 minutes to elute the immunoprecipitated complexes.

Purification of FBXL19-FS2 was performed using StrepTactin resin (IBA) as previously described with the exception of the wash buffer used (20 mM Tris pH 8.0, 150 mM NaCl, 0.2% NP-40, 1 mM DTT, 5% glycerol, cOmplete protease inhibitor cocktail) (Farcas et al., 2012). Between 10-15mg of nuclear extract was used for each large scale purification. The samples were treated with 75U/mL benzonase nuclease (Novagen) in order to disrupt nucleic acid-mediated interactions.

Mass spectrometry

Samples from FBXL19-FS2 affinity purifications were subjected to in-solution trypsin digestion and mass spectrometry analysis was performed as described previously (Farcas et al., 2012). Two biological replicates were performed. A control EV purification was included in each replicate. An interaction with identified proteins was only considered significant if absent from the EV data.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

609 **Antibodies**

610 Antibodies used for IP were rabbit anti-MED12 (A300-774A, Bethyl laboratories), rabbit anti-
611 CDK8 (A302-500A, Bethyl laboratories), rabbit anti-HA (3724, Cell Signaling Technology).
612 A polyclonal antibody against FBXL19 was prepared in-house by rabbit immunisation
613 (PTU/BS Scottish National Blood Transfusion Service) with a recombinant peptide encoding
614 for amino acids 137-336 of mouse FBXL19 protein. The FBXL19 peptide antigen was coupled
615 to Affigel 10 resin (BioRad) and the antibody was affinity-purified and concentrated.

616 Antibodies used for Western blot analysis were mouse anti-Flag M2 (F1804, Sigma-Aldrich),
617 mouse anti-SKP1 (sc-5281, Santa Cruz), goat anti-CDK8 (sc-1521, Santa Cruz), rabbit anti-
618 MED12 (A300-774A, Bethyl laboratories), rabbit anti-MED13L (A302-420A, Bethyl
619 laboratories), rabbit anti-MED13 (GTX129674, Genetex), rabbit anti-MED1, (A300-793A,
620 Bethyl laboratories), rabbit anti-MED26 (A302-370A, Bethyl laboratories), rabbit anti-T7
621 (13246, Cell Signaling Technology), rabbit anti-TBP (ab818, Abcam), rabbit anti-RNF20
622 (11974, Cell Signaling Technology), rabbit anti-ubiquityl-Histone H2B (Lys120) (5546, Cell
623 Signaling Technology), and mouse anti-ubiquityl-Histone H2B (Lys120) (05-1312, Millipore).

624

625 **Generation of T7-FBXL19 ES cell line**

626 As the FBXL19 antibody failed to work reliably for Western blot analysis, we generated an
627 *Fbxl19*$^{fl/fl}$ ES cell line in which the endogenous *Fbxl19* gene is tagged with a 3xT7-2xStrepII
628 tag by CRISPR/Cas9 knock-in (Supplementary Figure S2C). This allowed us to determine the
629 efficiency of the OHT treatment and endogenous IPs. The tag was synthesised by GeneArt
630 (ThermoFischer Scientific) and the targeting construct was generated by PCR amplification to
631 introduce roughly 150bp homology arms flanking the 3xT7-2xStrepII tag. The PCR product
632 was cloned into pGEM®-T Easy Vector (Promega). The pSpCas9(BB)-2A-Puro vector was

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

obtained by Addgene (#48139). A sgRNA was designed to overlap with the stop codon of the *Fbxl19* gene (http://crispr.mit.edu/) and cloned into the Cas9 vector as previously described (Ran et al., 2013). *Fbxl19*$^{fl/fl}$ ES cells were transiently co-transfected with 1µg of Cas9-sgRNA plasmid and 3.5µg of targeting construct using Lipofectamine 3000 (ThermoFischer Scientific). Successfully transfected ES cells were selected for 48 hours with 1µg/ml puromycin. Individual clones were screened by Western blot and genotyping PCR to identify homozygous targeted clones.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was performed as described previously (Farcas et al., 2012) with slight modifications. Cells were fixed for 45 min with 2mM DSG (Thermo scientific) in PBS and 12.5 min with 1% formaldehyde (methanol-free, Thermo scientific). Reactions were quenched by the addition of glycine to a final concentration of 125µM. After cell lysis and chromatin extraction, chromatin was sonicated using a BioRuptor Pico sonicator (Diagenode), followed by centrifugation at 16,000×*g* for 20min at 4°C. 200µg chromatin diluted in ChIP dilution buffer (1% Triton-X100, 1mM EDTA, 20mM Tris-HCl (pH 8.0), 150mM NaCl) was used per IP. Chromatin was precleared with protein A Dynabeads blocked with 0.2mg/ml BSA and 50µg/ml yeast tRNA and incubated with the respective antibodies overnight at 4°C. Antibody-bound chromatin was purified using blocked protein A Dynabeads for 3 hours at 4°C. ChIP washes were performed as described previously (Farcas et al., 2012). ChIP DNA was eluted in ChIP elution buffer (1% SDS, 100mM $NaHCO_3$) and reversed cross-linked overnight at 65°C with 200mM NaCl and RNase A (Sigma). The reverse cross-linked samples were treated with 20µg/ml Proteinase K and purified using ChIP DNA Clean&Concentrator™ kit (Zymo research).

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The antibodies used for ChIP experiments were rabbit anti-FS2 (Farcas et al., 2012), rabbit anti-CDK8 (A302-500A, Bethyl laboratories), rabbit anti-MED12 (A300-774A, Bethyl laboratories), mouse anti-MED4 (PRC-MED4-16, DSHB), rabbit anti-SKP1 (12248, Cell Signaling Technology), rabbit anti-KDM2A (Farcas et al., 2012), rabbit anti-KDM2B (Farcas et al., 2012).

ChIP-sequencing

All ChIP-seq experiments were carried out in biological duplicates. ChIP-seq libraries for FS2-FBXL19 ChIP were prepared using the NEBNext Fast DNA fragmentation and library preparation kit for Ion Torrent (NEB, E6285S) following manufacturer's instructions. Briefly, 30-40ng ChIP or input DNA material was used. Libraries were size-selected for 250bp fragments using 2% E-gel SizeSelect gel (Thermo scientific) and amplified with 6 PCR cycles or used without PCR amplification. Templates were generated with Ion PI™ Template OT2 200 kit v3 and Ion PI™ Sequencing 200 kit v3, or with Ion PI IC 200 kit (Thermo scientrific). Libraries were sequenced on the Ion Proton Sequencer using Ion PI™ chips v2 (Thermo scientific).

ChIP-seq libraries for CDK8 ChIP were prepared using the NEBNext Ultra DNA Library Prep Kit, and sequenced as 40bp paired-end reads on Illumina NextSeq500 platform using NextSeq® 500/550 (75 cycles).

Reverse transcription and gene expression analysis

Total RNA was isolated using TRIzol reagent (Thermo scientific) following manufacturer's instructions and cDNA was synthesised from 400ng RNA using random primers and ImProm-

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

680 II Reverse Transcription system kit (Promega). RT-qPCR was performed using SensiMix™
681 SYBR® mix (Bioline). Idh1 and Atp6IP1 genes were used as house-keeping controls.

682

683 **4sU-labelling of nascent RNA transcripts**

684 For labelling of nascent RNA transcripts, cells were grown in 15cm culture dishes. Labelling
685 was performed for 20 minutes at 37°C by the addition of 50mM 4-thiouridine (T4509, Sigma)
686 to the culture medium. After the incubation, the medium was removed and RNA was isolated
687 using TRIzol reagent (Thermo scientific) following manufacturer's instructions. The total
688 RNA was treated with TURBO DNA-free kit (Ambion, Thermo scientific) in order to remove
689 contaminating genomic DNA. 300µg RNA was biotinylated using 600µg Biotin-HPDP
690 (21341, Pierce, Thermo scientific) in Biotinylation buffer (100mM Tris-HCl pH 7.4, 10mM
691 EDTA). The reaction was carried out for one and half hour on a rotor at room temperature.
692 Unincorporated Biotin-HPDP was removed by chloroform/isoamylalcohol (24:1, Sigma) wash
693 followed by isopropanol precipitation of the biotinylated RNA. Labelled biotinylated RNA was
694 isolated using µMacs Streptavidin Kit (130-074-101, Miltenyi) following manufacturer's
695 instructions and purified using RNeasy MinElute Cleanup kit (Qiagen). The quality of the RNA
696 was confirmed using the RNA pico Bioanalyser kit (Agilent).

697 **4sU RNA sequencing**

698 Up to 1µg purified 4sU-labelled RNA was used to prepare libraries for 4sU-RNA-seq.
699 Ribosomal RNA was removed using NEBNext® rRNA Depletion Kit and libraries were
700 prepared using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina (NEB)
701 following manufacturer's instructions. Library quality was assessed using the High-sensitivity
702 DS DNA Bioanalyser kit (Agilent) and the concentration was measured by quantitative PCR
703 using KAPA Library quantification standards for Illumina (KAPA Biosystems). All 4sU RNA-

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

seq experiments were carried out in biological triplicates. 4sU RNA-seq libraries were sequenced as 80 paired-end reads on Illumina NextSeq500 platform using NextSeq® 500/550 High Output Kit v2 (150 cycles).

Analysis of high throughput sequencing

Sequencing reads were aligned to the mouse genome (mm10) using bowtie2 (Langmead and Salzberg, 2012) with the "--no-mixed" and "--no-discordant" options. Reads that mapped more than once to the genome were discarded. For 4sU RNA-seq analysis, rRNA reads were initially removed by aligning the data to mouse rRNA genomic sequences (GenBank: BK000964.3) using bowtie2. The rRNA-depleted reads were next aligned against mm10 genome using the STAR RNA-seq aligner (Dobin et al., 2013). To improve mapping of nascent, intronic sequences, a second alignment step with bowtie2 was included using the reads which failed to map using STAR. PCR duplicates were removed using samtools (Li et al., 2009). Biological replicates were randomly downsampled to the same number of reads for each individual replicate and merged for visualisation. Bigwig files were generated using MACS2 (Zhang et al., 2008) and visualised using the using the UCSC Genome Browser (Raney et al., 2014).

Peak calling was performed using the MACS2 function with the "-broad" option using the biological replicates with matched input (Zhang et al., 2008). Peaks mapping to a custom "blacklist" of artificially high genomic regions were discarded. Only peaks called in both replicates were considered. Differential analysis of CDK8 binding was done using DiffReps (Shen et al., 2013) and the called differential regions were overlapped with CDK8 peaks. Changes in binding of log2FC less or more than 0.5 with adjusted p-value below 0.05 were considered significant.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Differential expression analysis was performed using the DESeq2 package (Love et al., 2014), version 1.6.3, with R version 3.1.1. Counts were quantified using the summarizeOverlaps() function in R in the mode "Union" and inter.feature=FALSE. Genes with an adjusted p-value of below 0.1 and a fold change of at least 1.5 were considered differentially expressed. Statistical analysis was performed using two-sample Wilcoxon rank sum test. Gene ontology analysis was done using DAVID 6.8 (Huang da et al., 2009). For CDK8 differentially-bound peaks, all CDK8 peaks were provided as background. For misregulated genes analysis, all RefSeq genes were provided as background. ATAC peaks were obtained from (King and Klose, 2017).

Accession numbers

ChIP-seq and RNA-seq data from the present study are available to download at GSE98756. Previously published studies used for analysis include mouse ES cell BioCAP (GSE43512, (Long et al., 2013b)), Kdm2B ChIP-seq (GSE55698, (Blackledge et al., 2014)), Kdm2A ChIP-seq (GSE41267, (Farcas et al., 2012)), H3K27me3 ChIP-seq (GSE83135, (Rose et al., 2016)), H3K4me3 ChIP-seq (GSE93538, (Brown et al., 2017)).

Acknowledgments

We would like to thank Anne Turberfield for help with the 4sU-RNAseq protocol and useful discussion, and Hamish King for assistance with computational analysis and critical comments and suggestions. We thank David Brown for initiating the FBXL19 experiments, and Anne Turberfield and Neil Blackledge for critical reading of the manuscript. We would like to thank Ed Hookway and Udo Oppermann for sequencing support on the NextSeq500 at the Botnar sequencing facility in Oxford. Work in the Klose laboratory is supported by the Wellcome

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

751 Trust, the Lister Institute of Preventive Medicine and the European Research Council. T.K. and
752 H.K. are supported by the AMED-CREST programme from the Japan Agency for Medical
753 Research and Development. A.F. is supported by a Sir Henry Wellcome Postdoctoral
754 Fellowship.

755

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Adler, A.S., McCleland, M.L., Truong, T., Lau, S., Modrusan, Z., Soukup, T.M., Roose-Girma, M., Blackwood, E.M., and Firestein, R. (2012). CDK8 maintains tumor dedifferentiation and embryonic stem cell pluripotency. Cancer research *72*, 2129-2139.

Akoulitchev, S., Chuikov, S., and Reinberg, D. (2000). TFIIH is negatively regulated by cdk8-containing mediator complexes. Nature *407*, 102-106.

Alarcon, C., Zaromytidou, A.I., Xi, Q., Gao, S., Yu, J., Fujisawa, S., Barlas, A., Miller, A.N., Manova-Todorova, K., Macias, M.J., *et al.* (2009). Nuclear CDKs drive Smad transcriptional activation and turnover in BMP and TGF-beta pathways. Cell *139*, 757-769.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nature reviews. Molecular cell biology *16*, 155-166.

Audetat, K.A., Galbraith, M.D., Odell, A.T., Lee, T., Pandey, A., Espinosa, J.M., Dowell, R.D., and Taatjes, D.J. (2017). A Kinase-Independent Role for Cyclin-Dependent Kinase 19 in p53 Response. Molecular and cellular biology *37*.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell *86*, 263-274.

Bancerek, J., Poss, Z.C., Steinparzer, I., Sedlyarov, V., Pfaffenwimmer, T., Mikulic, I., Dolken, L., Strobl, B., Muller, M., Taatjes, D.J., *et al.* (2013). CDK8 kinase phosphorylates transcription factor STAT1 to selectively regulate the interferon response. Immunity *38*, 250-262.

Belakavadi, M., and Fondell, J.D. (2010). Cyclin-dependent kinase 8 positively cooperates with Mediator to promote thyroid hormone receptor-dependent transcriptional activation. Molecular and cellular biology *30*, 2437-2448.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Black, J.C., Choi, J.E., Lombardo, S.R., and Carey, M. (2006). A mechanism for coordinating chromatin modification and preinitiation complex assembly. Molecular cell *23*, 809-818.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., *et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Blackledge, N.P., Long, H.K., Zhou, J.C., Kriaucionis, S., Patient, R., and Klose, R.J. (2012). Bio-CAP: a versatile and highly sensitive technique to purify and characterise regions of non-methylated DNA. Nucleic acids research *40*, e32.

Blackledge, N.P., Thomson, J.P., and Skene, P.J. (2013). CpG island chromatin is shaped by recruitment of ZF-CxxC proteins. Cold Spring Harbor perspectives in biology *5*, a018648.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Molecular cell *38*, 179-190.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

797 Blazek, E., Mittler, G., and Meisterernst, M. (2005). The mediator of RNA polymerase II.
798 Chromosoma *113*, 399-408.

799 Boyer, L.A., Mathur, D., and Jaenisch, R. (2006). Molecular control of pluripotency. Current
800 opinion in genetics & development *16*, 455-462.

801 Brown, D.A., Di Cerbo, V., Feldmann, A., Ahn, J., Ito, S., Blackledge, N.P., Nakayama, M.,
802 McClellan, M., Dimitrova, E., Turberfield, A.H., *et al.* (2017). The SET1 Complex Selects
803 Actively Transcribed Target Genes via Multivalent Interaction with CpG Island Chromatin.
804 Cell Rep *20*, 2313-2327.

805 Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J.,
806 Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs
807 deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell
808 *123*, 581-592.

809 Chi, Y., Huddleston, M.J., Zhang, X., Young, R.A., Annan, R.S., Carr, S.A., and Deshaies,
810 R.J. (2001). Negative regulation of Gcn4 and Msn2 transcription factors by Srb10 cyclin-
811 dependent kinase. Genes & development *15*, 1078-1092.

812 D'Urso, A., Takahashi, Y.H., Xiong, B., Marone, J., Coukos, R., Randise-Hinchliff, C., Wang,
813 J.P., Shilatifard, A., and Brickner, J.H. (2016). Set1/COMPASS and Mediator are repurposed
814 to promote epigenetic transcriptional memory. eLife *5*.

815 Davis, M.A., Larimore, E.A., Fissel, B.M., Swanger, J., Taatjes, D.J., and Clurman, B.E.
816 (2013). The SCF-Fbw7 ubiquitin ligase degrades MED13 and MED13L and regulates CDK8
817 module association with Mediator. Genes & development *27*, 151-156.

818 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson,
819 M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*,
820 15-21.

821 Dong, S., Zhao, J., Wei, J., Bowser, R.K., Khoo, A., Liu, Z., Luketich, J.D., Pennathur, A., Ma,
822 H., and Zhao, Y. (2014). F-box protein complex FBXL19 regulates TGFbeta1-induced E-
823 cadherin down-regulation by mediating Rac3 ubiquitination and degradation. Molecular cancer
824 *13*, 76.

825 Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive
826 regulator of transcriptional elongation within the serum response network. Nature structural &
827 molecular biology *17*, 194-201.

828 Donner, A.J., Szostek, S., Hoover, J.M., and Espinosa, J.M. (2007). CDK8 is a stimulus-
829 specific positive coregulator of p53 target genes. Molecular cell *27*, 121-133.

830 Elmlund, H., Baraznenok, V., Lindahl, M., Samuelsen, C.O., Koeck, P.J., Holmberg, S.,
831 Hebert, H., and Gustafsson, C.M. (2006). The cyclin-dependent kinase 8 module sterically
832 blocks Mediator interactions with RNA polymerase II. Proc Natl Acad Sci U S A *103*, 15788-
833 15793.

834 Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S.,
835 Sims, D., Cerase, A., Sheahan, T.W., *et al.* (2012). KDM2B links the Polycomb Repressive
836 Complex 1 (PRC1) to recognition of CpG islands. eLife *1*, e00205.

837 Firestein, R., Bass, A.J., Kim, S.Y., Dunn, I.F., Silver, S.J., Guney, I., Freed, E., Ligon, A.H.,
838 Vena, N., Ogino, S., *et al.* (2008). CDK8 is a colorectal cancer oncogene that regulates beta-
839 catenin activity. Nature *455*, 547-551.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fondell, J.D., Ge, H., and Roeder, R.G. (1996). Ligand induction of a transcriptionally active thyroid hormone receptor coactivator complex. Proc Natl Acad Sci U S A *93*, 8329-8333.

Fryer, C.J., White, J.B., and Jones, K.A. (2004). Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. Molecular cell *16*, 509-520.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D., *et al.* (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327-1339.

He, J., Kallin, E.M., Tsukada, Y., and Zhang, Y. (2008). The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nature structural & molecular biology *15*, 1169-1175.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nature cell biology *15*, 373-384.

Hengartner, C.J., Myer, V.E., Liao, S.M., Wilson, C.J., Koh, S.S., and Young, R.A. (1998). Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. Molecular cell *2*, 43-53.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS letters *583*, 1713-1720.

Jeronimo, C., Langelier, M.F., Bataille, A.R., Pascal, J.M., Pugh, B.F., and Robert, F. (2016). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. Molecular cell *64*, 455-466.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Katoh, M., and Katoh, M. (2004). Identification and characterization of FBXL19 gene in silico. International journal of molecular medicine *14*, 1109-1114.

Keogh, M.C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., *et al.* (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell *123*, 593-605.

King, H.W., and Klose, R.J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. eLife *6*.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences *31*, 89-97.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. PLoS genetics *9*, e1003717.

Knuesel, M.T., Meyer, K.D., Bernecky, C., and Taatjes, D.J. (2009). The human CDK8 subcomplex is a molecular switch that controls Mediator coactivator function. Genes & development *23*, 439-451.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Lee, B.K., Lee, J., Shen, W., Rhee, C., Chung, H., and Kim, J. (2017). Fbxl19 recruitment to CpG islands is required for Rnf20-mediated H2B mono-ubiquitination. Nucleic acids research.

Lee, J.H., Voo, K.S., and Skalnik, D.G. (2001). Identification and characterization of the DNA binding domain of CpG-binding protein. The Journal of biological chemistry *276*, 44669-44676.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, Y., Ranish, J.A., Aebersold, R., and Hahn, S. (2001). Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes. The Journal of biological chemistry *276*, 7169-7175.

Long, H.K., Blackledge, N.P., and Klose, R.J. (2013a). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochemical Society transactions *41*, 727-740.

Long, H.K., Sims, D., Heger, A., Blackledge, N.P., Kutter, C., Wright, M.L., Grutzner, F., Odom, D.T., Patient, R., Ponting, C.P., *et al.* (2013b). Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. eLife *2*, e00348.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet *11*, 761-772.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mittler, G., Kremmer, E., Timmers, H.T., and Meisterernst, M. (2001). Novel critical role of a human Mediator complex for basal RNA polymerase II transcription. EMBO Rep *2*, 808-813.

Morris, E.J., Ji, J.Y., Yang, F., Di Stefano, L., Herr, A., Moon, N.S., Kwon, E.J., Haigis, K.M., Naar, A.M., and Dyson, N.J. (2008). E2F1 represses beta-catenin transcription and is antagonized by both pRB and CDK8. Nature *455*, 552-556.

Naar, A.M., Taatjes, D.J., Zhai, W., Nogales, E., and Tjian, R. (2002). Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. Genes & development *16*, 1339-1344.

Paoletti, A.C., Parmely, T.J., Tomomori-Sato, C., Sato, S., Zhu, D., Conaway, R.C., Conaway, J.W., Florens, L., and Washburn, M.P. (2006). Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. Proc Natl Acad Sci U S A *103*, 18928-18933.

Pavri, R., Lewis, B., Kim, T.K., Dilworth, F.J., Erdjument-Bromage, H., Tempst, P., de Murcia, G., Evans, R., Chambon, P., and Reinberg, D. (2005). PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator. Molecular cell *18*, 83-96.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

926 Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L.,
927 Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and plasticity
928 of repressive histone methylation states in mammalian chromatin. Molecular cell *12*, 1577-
929 1589.

930 Petrenko, N., Jin, Y., Wong, K.H., and Struhl, K. (2016). Mediator Undergoes a Compositional
931 Change during Transcriptional Activation. Molecular cell *64*, 443-454.

932 Poss, Z.C., Ebmeier, C.C., and Taatjes, D.J. (2013). The Mediator complex and transcription
933 regulation. Crit Rev Biochem Mol Biol *48*, 575-608.

934 Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome
935 engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.

936 Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N.,
937 Paten, B., Zweig, A.S., Karolchik, D., *et al.* (2014). Track data hubs enable visualization of
938 user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics *30*,
939 1003-1005.

940 Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler,
941 B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based
942 communication between Polycomb repressive complexes. eLife *5*.

943 Sato, S., Tomomori-Sato, C., Parmely, T.J., Florens, L., Zybailov, B., Swanson, S.K., Banks,
944 C.A., Jin, J., Cai, Y., Washburn, M.P., *et al.* (2004). A set of consensus mammalian mediator
945 subunits identified by multidimensional protein identification technology. Molecular cell *14*,
946 685-691.

947 Schubeler, D. (2015). Function and information content of DNA methylation. Nature *517*, 321-
948 326.

949 Shen, L., Shao, N.Y., Liu, X., Maze, I., Feng, J., and Nestler, E.J. (2013). diffReps: detecting
950 differential chromatin modification sites from ChIP-seq data with biological replicates. PLoS
951 One *8*, e65598.

952 Skaar, J.R., Pagan, J.K., and Pagano, M. (2014). SCF ubiquitin ligase-targeted therapies. Nat
953 Rev Drug Discov *13*, 889-903.

954 Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. (1997). F-box proteins are
955 receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. Cell *91*,
956 209-219.

957 Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to
958 developmental control. Nat Rev Genet *13*, 613-626.

959 Taatjes, D.J., Naar, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function,
960 and activator-induced conformations of the CRSP coactivator. Science *295*, 1058-1062.

961 Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz,
962 H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., *et al.* (2011). Human mediator subunit
963 MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.

964 Tan, M.K., Lim, H.J., Bennett, E.J., Shi, Y., and Harper, J.W. (2013). Parallel SCF adaptor
965 capture proteomics reveals a role for SCFFBXL17 in NRF2 activation via BACH1 repressor
966 turnover. Molecular cell *52*, 9-24.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

967 Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton,
968 A., Andrews, R., James, K.D., *et al.* (2010). CpG islands influence chromatin structure via the
969 CpG-binding protein Cfp1. Nature *464*, 1082-1086.

970 Tsai, K.L., Sato, S., Tomomori-Sato, C., Conaway, R.C., Conaway, J.W., and Asturias, F.J.
971 (2013). A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA
972 polymerase II interaction. Nature structural & molecular biology *20*, 611-619.

973 Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and
974 Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins.
975 Nature *439*, 811-816.

976 Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a
977 mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC
978 domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain
979 protein 1. Molecular and cellular biology *20*, 2108-2121.

980 Wei, J., Mialki, R.K., Dong, S., Khoo, A., Mallampalli, R.K., Zhao, Y., and Zhao, J. (2013). A
981 new mechanism of RhoA ubiquitination and degradation: roles of SCF(FBXL19) E3 ligase and
982 Erk2. Biochimica et biophysica acta *1833*, 2757-2764.

983 Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B.,
984 Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-
985 enhancers at key cell identity genes. Cell *153*, 307-319.

986 Wojcik, C., Tanaka, K., Paweletz, N., Naab, U., and Wilk, S. (1998). Proteasome activator
987 (PA28) subunits, alpha, beta and gamma (Ki antigen) in NT2 neuronal precursor cells and
988 HeLa S3 cells. European journal of cell biology *77*, 151-160.

989 Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive
990 complex 1 to CpG islands and regulates H2A ubiquitylation. Molecular cell *49*, 1134-1146.

991 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C.,
992 Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS).
993 Genome Biol *9*, R137.

994 Zhao, J., Mialki, R.K., Wei, J., Coon, T.A., Zou, C., Chen, B.B., Mallampalli, R.K., and Zhao,
995 Y. (2013). SCF E3 ligase F-box protein complex SCF(FBXL19) regulates cell migration by
996 mediating Rac1 ubiquitination and degradation. FASEB journal : official publication of the
997 Federation of American Societies for Experimental Biology *27*, 2611-2619.

998 Zhao, J., Wei, J., Mialki, R.K., Mallampalli, D.F., Chen, B.B., Coon, T., Zou, C., Mallampalli,
999 R.K., and Zhao, Y. (2012). F-box protein FBXL19-mediated ubiquitination and degradation of
1000 the receptor for IL-33 limits pulmonary inflammation. Nature immunology *13*, 651-658.

1001

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

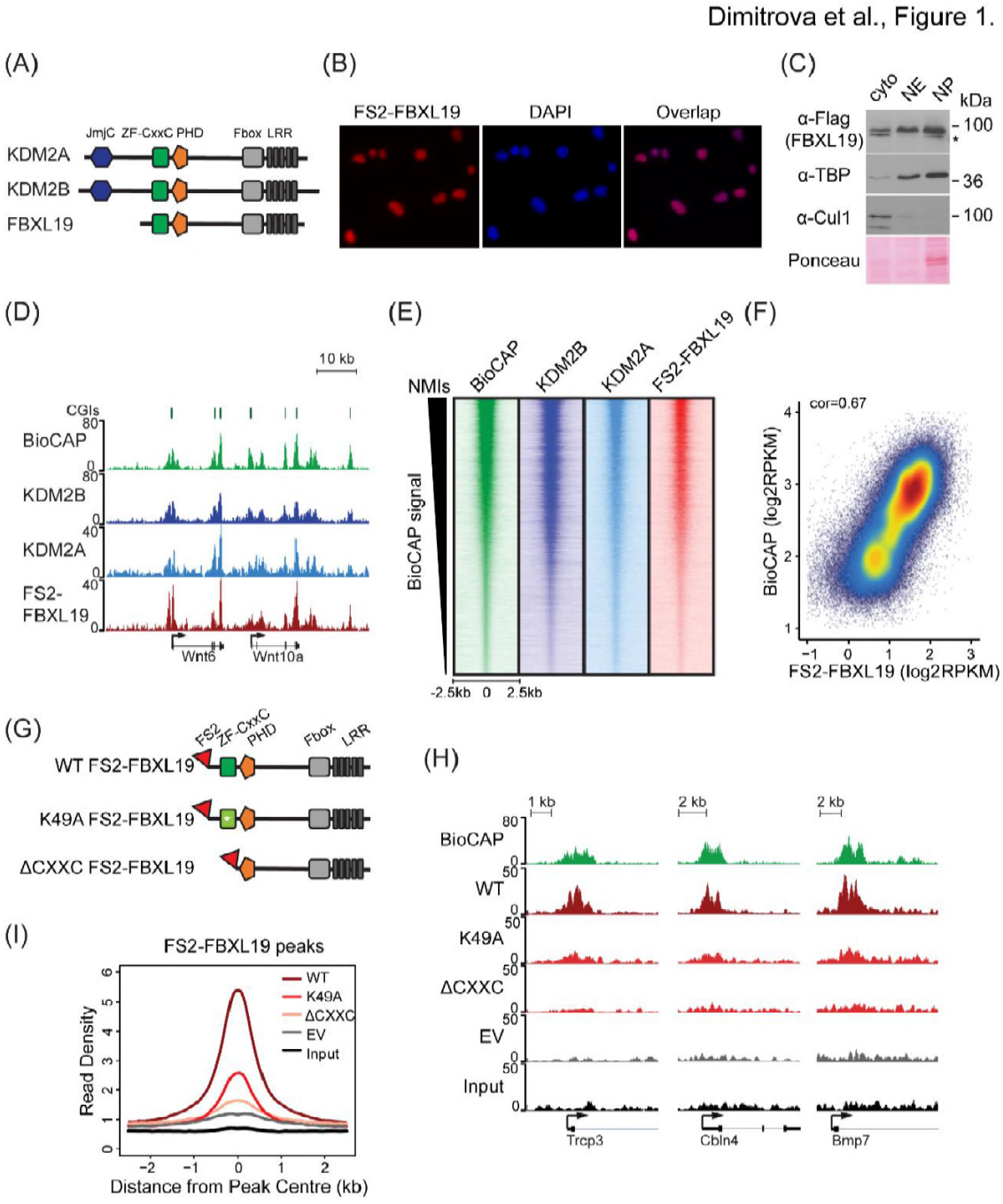


Figure 1: FBXL19 binds CpG islands genome-wide via its ZF-CxxC domain.

(A) A schematic illustrating KDM2 protein domain architecture.

(B) Immuno-fluorescent staining for FS2-FBXL19 in ES cells.

(C) ES cell fractionation and Western blot for factors enriched in the nucleus (TBP), the cytoplasm (CUL1), and FS2-FBXL19. Ponceau S staining indicates protein loading. Cyto – cytoplasmic fraction, NE – soluble nuclear extract, NP – insoluble nuclear pellet. The asterisk indicates a non-specific band.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(D) Screen shots showing ChIP-seq traces for KDM2B, KDM2A and FS2-FBXL19. BioCAP is included to indicate the location of non-methylated islands (NMIs) and computationally predicted CGIs are illustrated above.

(E) Heatmaps showing enrichment of KDM2 proteins over a 5kb region centred on NMIs (n=27698), sorted by decreasing BioCAP signal. BioCAP is shown for comparison.

(F) A scatter plot showing the Spearman correlation between FS2-FBXL19 ChIP-seq signal and BioCAP signal at NMIs.

(G) A schematic illustrating the FS2-FBXL19 transgenes.

(H) A screen shot showing ChIP-seq traces for WT FS2-FBXL19 and ZF-CxxC FS2-FBXL19 mutants as in G. EV indicates empty vector control.

(I) A metaplot analysis of WT FS2-FBXL19 and ZF-CxxC FS2-FBXL19 mutants over FBXL19 peaks.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

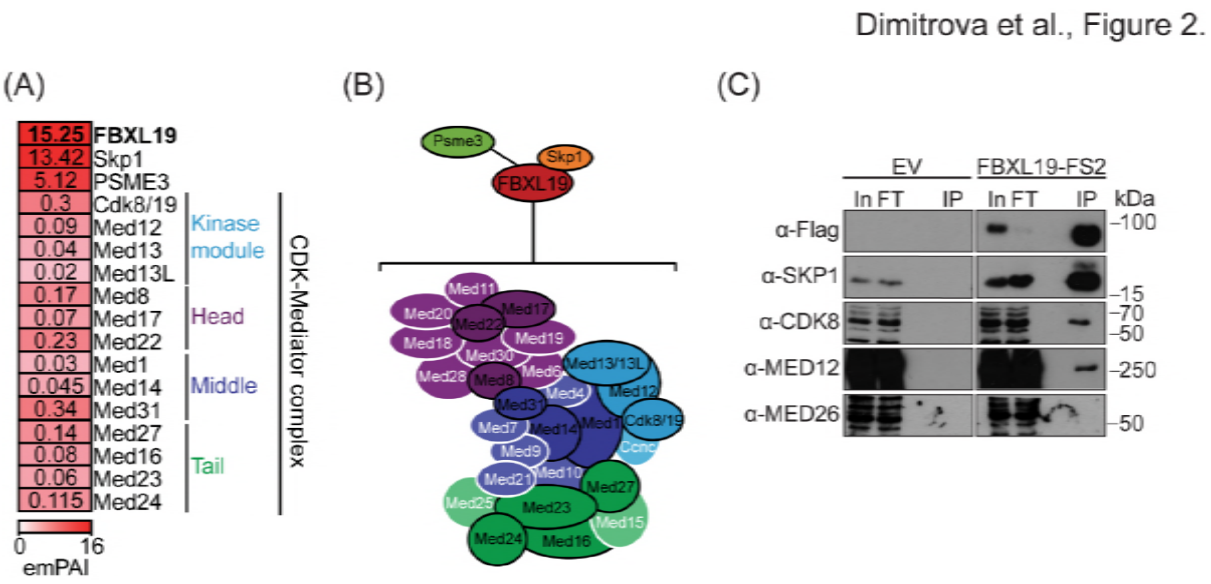


Figure 2: FBXL19 interacts with CDK-Mediator complex in ES cells.

(A) A heatmap representing the empirically modified protein abundance index (emPAI) of identified FBXL19-interacting proteins by affinity purification mass spectrometry (AP-MS). Data shown is an average of two biological replicates. The location of the identified Mediator components within the holocomplex is summarised on the right and in (B).

(B) A schematic representation of the identified FBXL19-interacting proteins. Subunits of the CDK-Mediator complex not identified by AP-MS are shown in white.

(C) Western blot analysis of FS2-FBXL19 and control purifications (EV) probed with the indicated antibodies.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova et al. Figure 3.

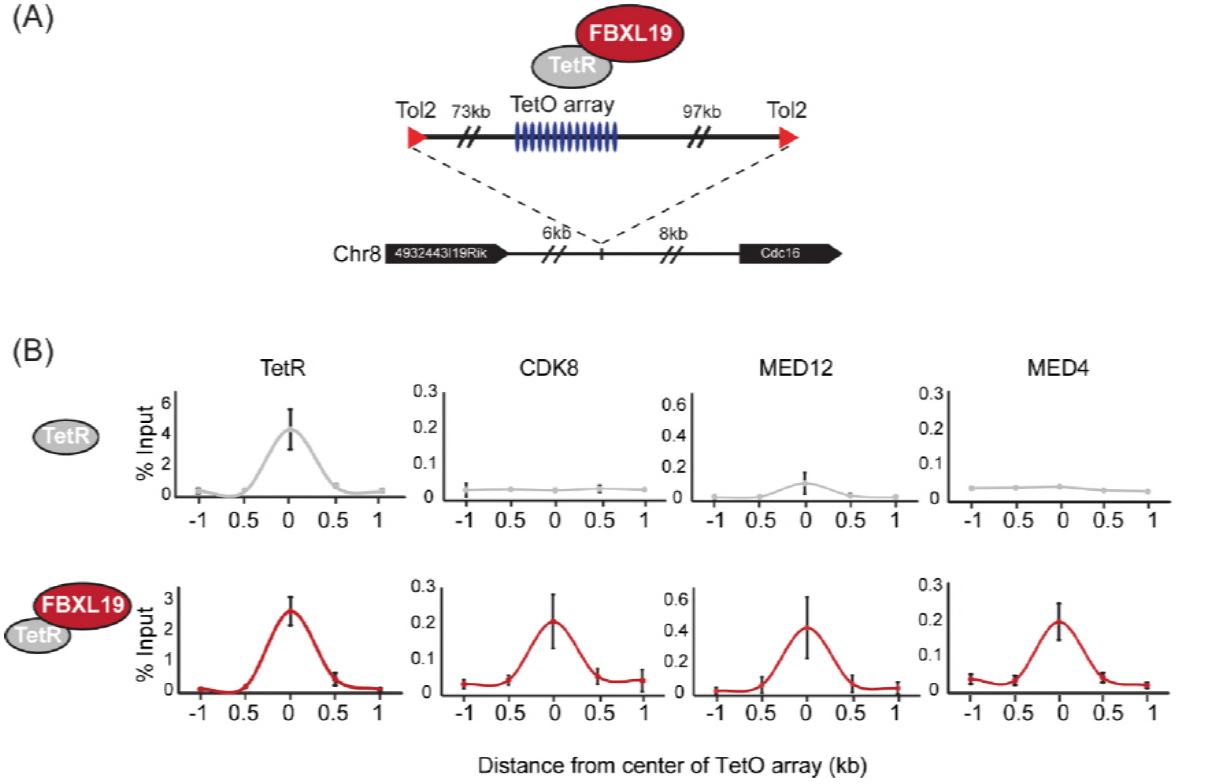


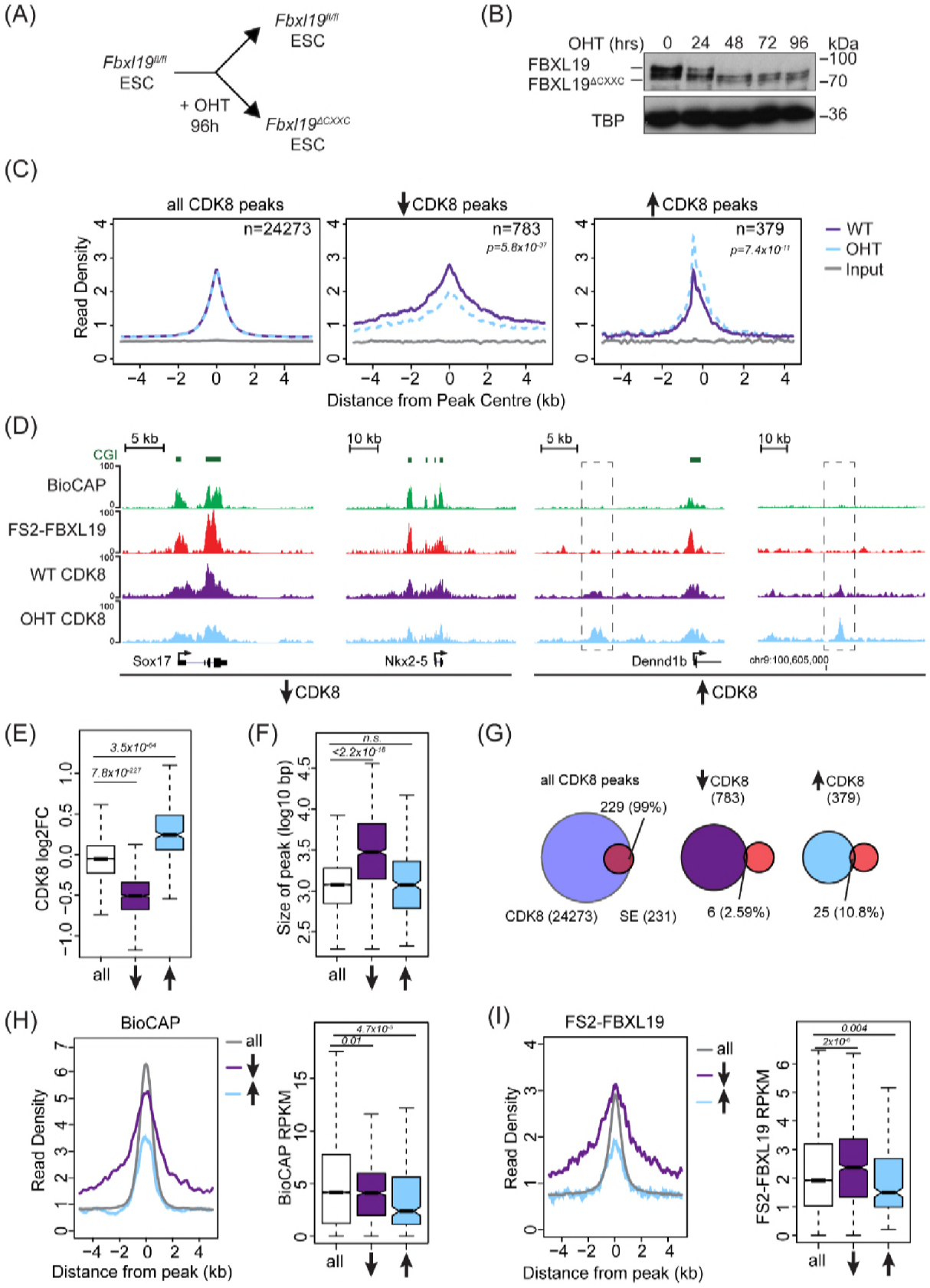
Figure 3: FBXL19 recruits CDK-Mediator to chromatin.

(A) A schematic representation of the integration site of the TetO array.

(B) ChIP-qPCR analysis of the binding of the indicated proteins across the TetO array in TetR only (top) and TetR-FBXL19 (bottom) ES cell lines. The x-axis indicates the spatial arrangement of the qPCR amplicons with respect to the center of the TetO array (in kb). Error bars represent SEM of three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova et al. Figure 4.



1042

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4: FBXL19 is required for appropriate CDK8 occupancy at a subset of CpG island promoters.

(A) A schematic illustrating how addition of OHT (4-hydroxytamoxifen) leads to the generation of *Fbxl19*$^{\Delta CXXC}$ ES cells.

(B) Western blot analysis of an OHT-treatment time course in the *Fbxl19*$^{fl/fl}$ ES cell line. Hours of treatment are shown. FBXL19 protein was C-terminally tagged with a T7 epitope (Figure S2B and S2C) to allow for Western blot detection with an anti-T7 antibody. TBP was probed as a loading control.

(C) Metaplots showing CDK8 enrichment in *Fbxl19*$^{fl/fl}$ ES cells (WT) and *Fbxl19*$^{\Delta CXXC}$ ES cells (OHT) at all CDK8 peaks (left), peaks showing decreased CDK8 occupancy (↑, middle) and peaks showing increased CDK8 occupancy (↓, right). The number of total peaks in each group is indicated. p-values denote statistical significance calculated by Wilcoxon rank sum test comparing ChIP-seq read counts between WT and OHT samples across a 1.5-kbp interval flanking the center of CDK8 peaks.

(D) Screen shots showing ChIP-seq traces for CDK8 in wild type *Fbxl19*$^{fl/fl}$ ES cells (WT) and *Fbxl19*$^{\Delta CXXC}$ ES cells (OHT). BioCAP and FS2-FBXL19 tracks are given for comparison. Left – CDK8 peaks showing reduced CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells; Right – CDK8 peaks showing increased CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells (indicated with rectangles).

(E) Boxplots showing log2 fold change (log2FC) of CDK8 ChIP-seq signal (RPKM) at all CDK8 peaks (n=24273), and those with reduced CDK8 (n=783, ↓) and increased CDK8 (n=379, ↑). p-values were calculated using a Wilcoxon rank sum test.

(F) Boxplots showing size of CDK8 peaks as in (E). p-values were calculated using a Wilcoxon rank sum test.

(G) Venn diagrams representing the overlap between CDK8 peaks and super enhancers (SE) at CDK8 peaks as in (E). Percent overlap of all SEs is indicated.

(H) A metaplot (left) showing BioCAP enrichment at CDK8 peaks as in (E) and boxplot quantification (right) of BioCAP RPKM levels. p-values calculated using Wilcoxon rank sum test are indicated.

(I) A metalplot (left) showing FS2-FBXL19 enrichment at CDK8 peaks as in (E) and boxplot quantification (right) of FS2-FBXL19 RPKM levels. p-values calculated using Wilcoxon rank sum test are indicated.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova el al. Figure 5.

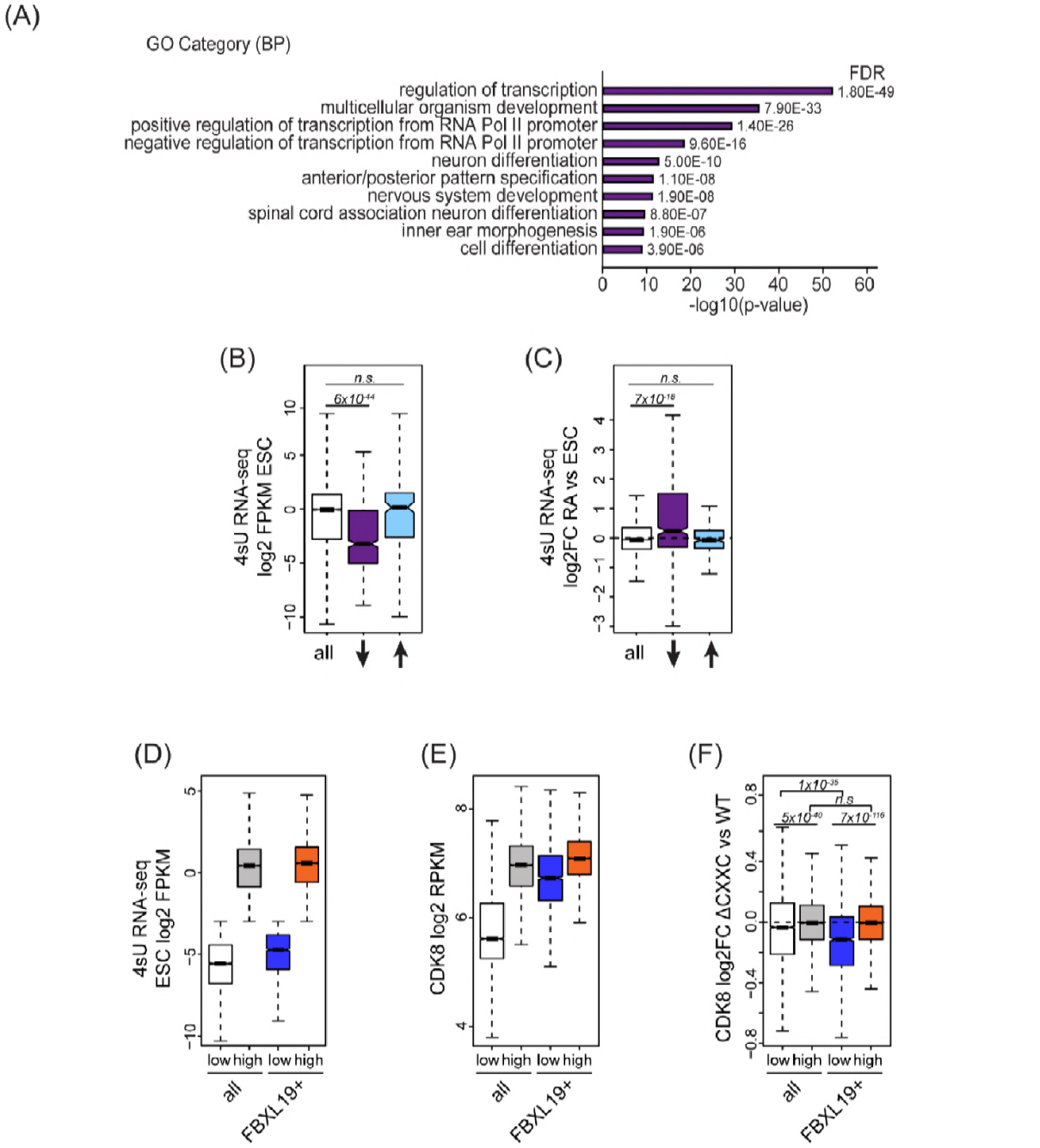


Figure 5. FBXL19 targets CDK8 to promoters of silent developmental genes in ES cells.

(A) Gene ontology analysis of genes associated with a decrease in CDK8 binding.

(B) A boxplot showing expression levels (log2FPKM) of CDK8-bound genes in wild type ES cells. Genes are divided based on CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells as in Figure 4E. p-values were calculated using a Wilcoxon rank sum test.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(C) A boxplot showing the change in gene expression (log2 fold change) observed by 4sU RNA-seq of CDK8-bound genes (as in Figure 4E) following RA treatment. p-values were calculated using a Wilcoxon rank sum test.

(D) A boxplot comparing gene expression levels (log2FPKM) of all (n=19310) and FBXL19-bound (FBXL19+, n=11368) genes separated by low (all genes n=7417; FBXl19+ genes n=2031) and high expression levels (all genes n=11893, FBXl19+ genes n=9337) in ES cells (based on Figure S5B).

(E) A boxplot showing CDK8 enrichment at all and FBXL19-bound genes separated by expression level as in (D).

(F) A boxplot showing change in CDK8 binding at the TSSs of all and FBXL19-bound genes divided by expression level as in (D). p-values were calculated using a Wilcoxon rank sum test.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

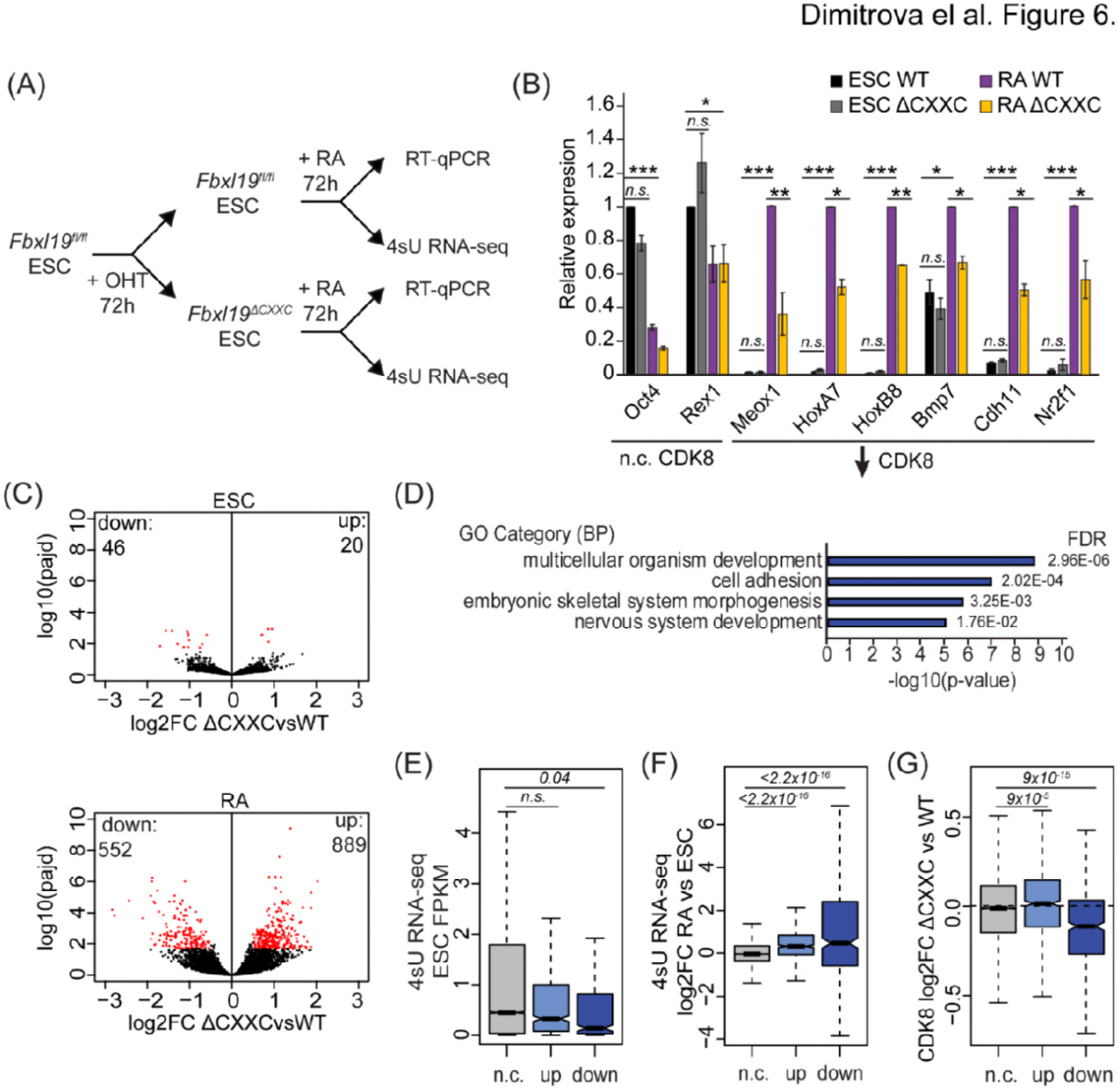


Figure 6: Removing the CpG island-binding domain of FBXL19 results in a failure to induce developmental genes during ES cell differentiation.

(A) A schematic illustrating the OHT treatment and differentiation approach.

(B) RT-qPCR gene expression analysis of genes showing decreased CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells before (ESC) and after RA induction. Expression is relative to the average of two house-keeping genes. Error bars show SEM of three biological replicates, asterisk represent statistical significance calculated by Student T-test: * p<0.05, ** p<0.01, *** p<0.001.

(C) Volcano plots showing differential expression (log2 fold change) comparing WT and *Fbxl19*$^{\Delta CXXC}$ cells in the ES cell (top) and RA-induced state (bottom). Differentially expressed

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes (log2FC < -0.5 or log2FC > 0.5, padj <0.1) are shown in red. The number of genes considered to be significantly altered in expression are indicated.

(D) Gene ontology analysis of genes with decreased expression in *Fbxl19*$^{\Delta CXXC}$ ES cells following RA treatment.

(E) A boxplot indicating the expression level (FPKM) based on 4sU RNA-seq in wild type ES cells of the gene groupings in (C): n.c. genes n=17869 (no change), up genes = 889, down genes = 552. p-values calculated using a Wilcoxon rank sum test are indicated.

(F) A boxplot indicating the log2 fold change in gene expression of the gene groupings in (E) upon RA differentiation of ES cells. p-values calculated using a Wilcoxon rank sum test are indicated.

(G) A boxplot indicating the change in CDK8 binding (log2FC) at the promoters of the gene groupings in (E) in ES cells. p-values calculated using a Wilcoxon rank sum test are indicated.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

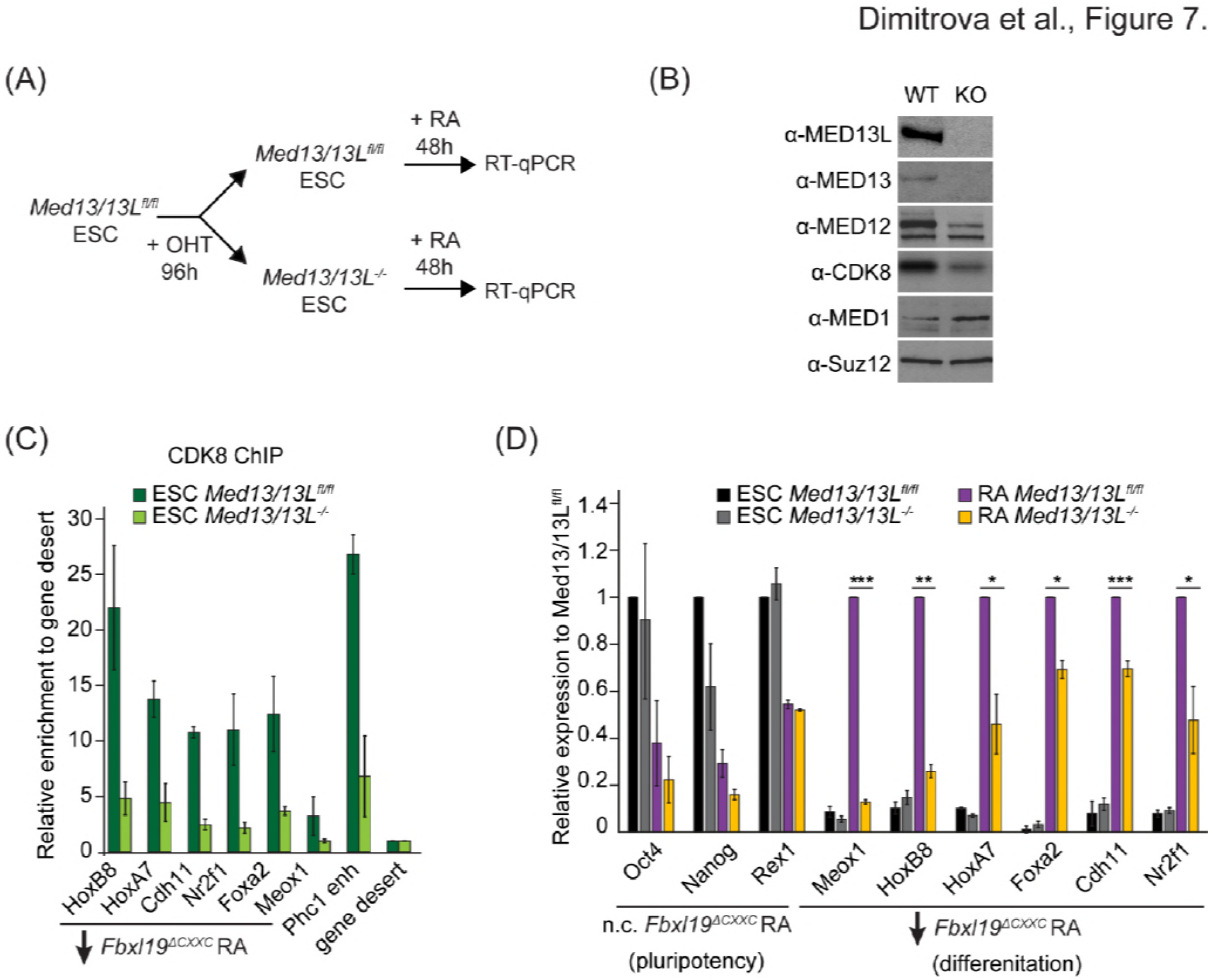


Figure 7: FBXL19 target genes rely on CDK-Mediator for activation during differentiation.

(A) A schematic illustrating the OHT treatment to generate MED13/13L KO ES cells and differentiation approach.

(B) Western blot analysis showing the efficiency of MED13/13 knock-out upon 96h OHT treatment of *Med13/13L*$^{fl/fl}$ ES cells. Suz12 was blotted as a loading control.

(C) ChIP-qPCR showing CDK8 enrichment in WT and MED13/13L KO ES cells. Enrichment is relative to gene desert control region. Error bars show standard deviation of two biological replicates.

(D) RT-qPCR gene expression analysis in WT and MED13/13L KO ES cells before and after RA induction. Expression was normalised to the expression of the PolIII-transcribed gene *tRNA-Lys* and is represented as relative to WT ES cells (for pluripotency genes) or RA-treated WT cells (for differentiation markers). Error bars show SEM of three biological replicates, asterisks represent statistical significance calculated by Student T-test: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

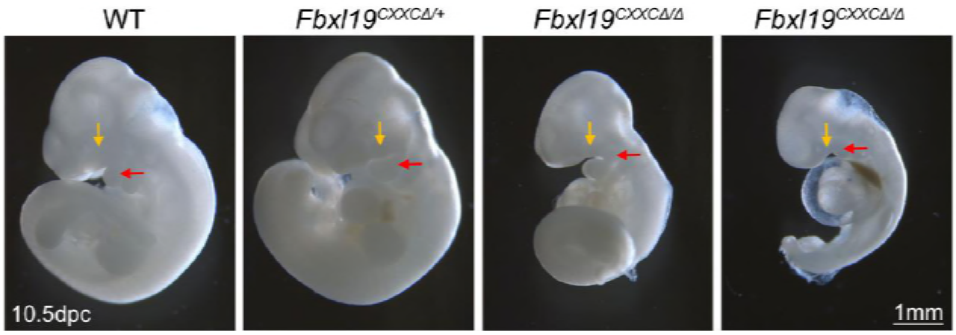
bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova el al. Figure 8.

(A)

Stage	WT	*Fbxl19*$^{CXXC\Delta/+}$	*Fbxl19*$^{CXXC\Delta/\Delta}$	Total
9.5 dpc	4 (21%)	9 (47%)	6 (32%)	19
10.5 dpc	20 (26%)	40 (52%)	17 (22%)	77
newborn	8 (33%)	16 (67%)	0 (0%)	24

(B)



1135

1136 **Figure 8: Deletion of the ZF-CxxC domain of FBXL19 leads to mouse embryonic lethality**

1137 (A) After crossing heterozygotes, the number of live embryos (9.5 dpc or 10.5 dpc) or newborn
1138 pups with the indicated genotypes is indicated with the percentages shown in parentheses.
1139 (B) The morphological changes in 10.5 dpc *Fbxl19*$^{CXXC\Delta/\Delta}$ embryos are indicated. Lateral views
1140 of wild-type (wt), together with heterozygous (*Fbxl19*$^{CXXC\Delta/+}$) and homozygous (*Fbxl19*$^{CXXC\Delta/\Delta}$)
1141 mutants are shown. Maxillary and mandibular components of first branchial arches are
1142 indicated by yellow and red arrows, respectively.

1143

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

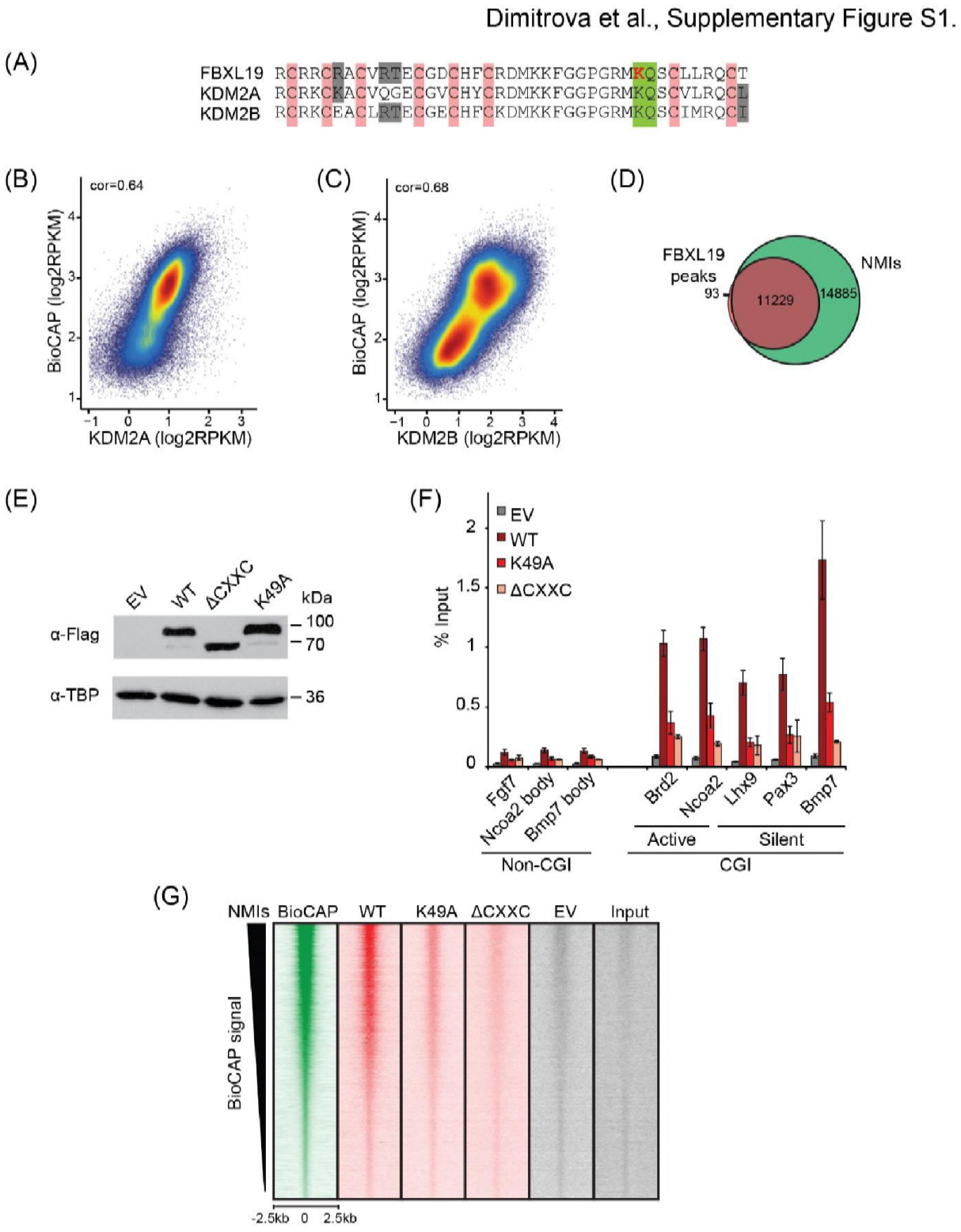


Figure S1: FBXL19 binds CpG islands genome-wide via its ZF-CxxC domain.

(A) Alignment of the KDM2 protein ZF-CxxC domains. Pink indicates the location of cysteine residues that coordinate Zn^{2+} binding in the ZF-CxxC domain. The conserved lysine K49 within the CpG-binding KQ motif (green) in FBXL19 is indicated in red.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(B-C) A scatterplot showing the Spearman correlation between KDM2A or KDM2B and BioCAP at NMIs.

(D) A Venn diagram showing the overlap between FS2-FBXL19 peaks (n=11322) and NMIs (n=27698).

(E) Western blot analysis showing expression of wild-type (WT) FS2-FBXL19, ΔCxxC FS2-FBXL19 and ZF-CxxC-mutant (K49A) FS2-FBXL19 from nuclear extracts of the respective stable ES cell lines.

(F) ChIP-qPCR for WT and mutant FS2-FBXL19 enrichment at gene promoters. Error bars show SEM of three biological replicates.

(G) Heatmap analysis at NMI peaks (n=27698), showing ChIP-seq data for BioCAP, WT FS2-FBXL19 and ZF-CxxC mutants over a 5kb region centred on the NMI, sorted by decreasing BioCAP signal.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

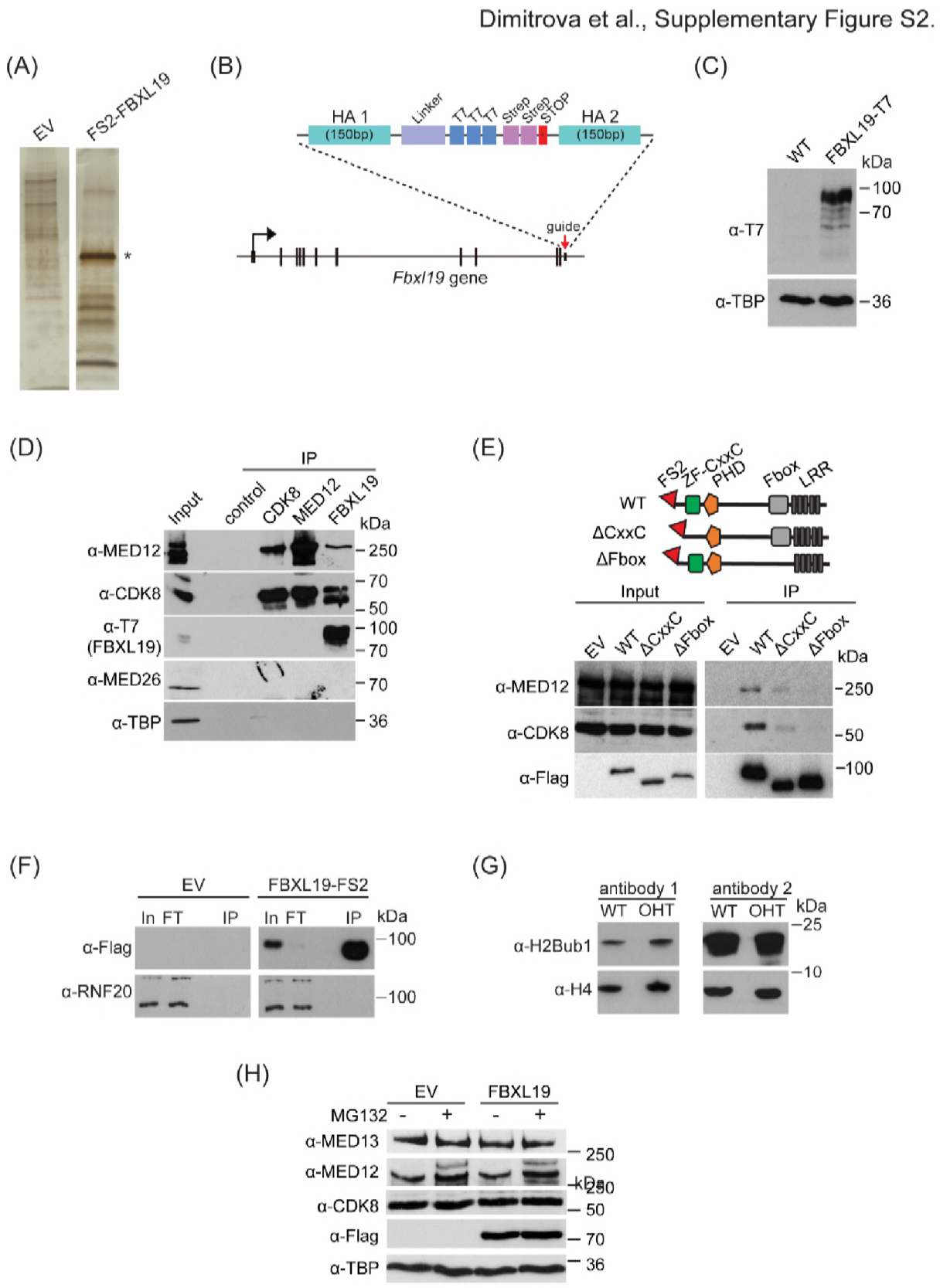
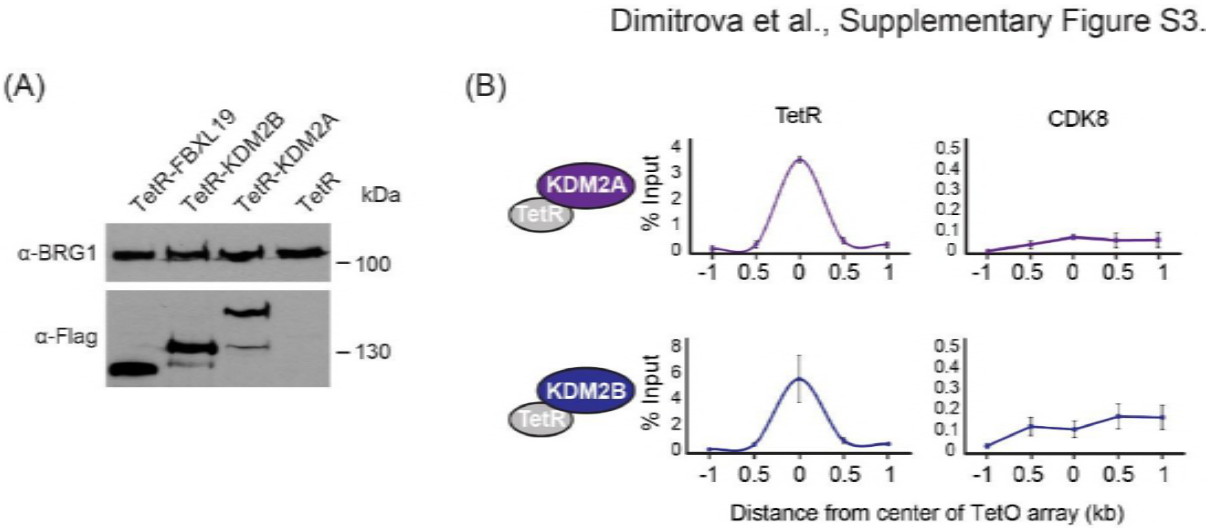


Figure S2: FBXL19 interacts with CDK-Mediator complex in ES cells.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(A) A representative silver stained gel for FS2-FBXL19 purification and an empty vector control (EV) purification. Asterisk identifies the band corresponding to FS2-FBXL19 protein.
(B) In order to visualise FBXL19 protein by western blot in Figure S2D and Figure 4B, the Fbxl19 gene was tagged by T7 knock-in in *Fbxl19*$^{fl/fl}$ ES cells using the CRISPR Cas9 system. A schematic representation of the generation of the C-terminal T7 knock-in Fbxl19 is shown. HA1/2 indicate the homology arms of the targeting construct.
(C) Western blot analysis of the expression of T7-FBXL19 from nuclear extract of the generated T7 knock-in ES cell line.
(D) Western blot analysis of endogenous co-immunoprecipitation (IP) of FBXL19, CDK8 and MED12 from ES cell nuclear extracts. A control IP using a non-specific antibody (α-HA) was included.
(E) A schematic illustration of the different FS2-FBXL19 truncation mutants and Western blot analysis of purification of FS2-FBXL19 mutants from HEK293T cells probed with the indicated antibodies.
(F) Western blot analysis of FS2-FBXL19 and control purifications (EV) probed with the indicated antibodies.
(G) Western blot analysis of histone extracts generated from *Fbxl19*$^{fl/fl}$ (WT) and *Fbxl19*$^{\Delta CXXC}$ (OHT) ES cells probed with two different antibodies recognizing ubiquitylated H2B K120 (H2Bub1). H4 was used as a loading control.
(H) Western blot analysis of nuclear extracts from HEK293T cells transiently transfected with empty vector (EV) or Flag-FBXL19-expressing vector without (-) or following MG132 treatment (+). Blots were probed with the indicated antibodies. TBP was used as loading control.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



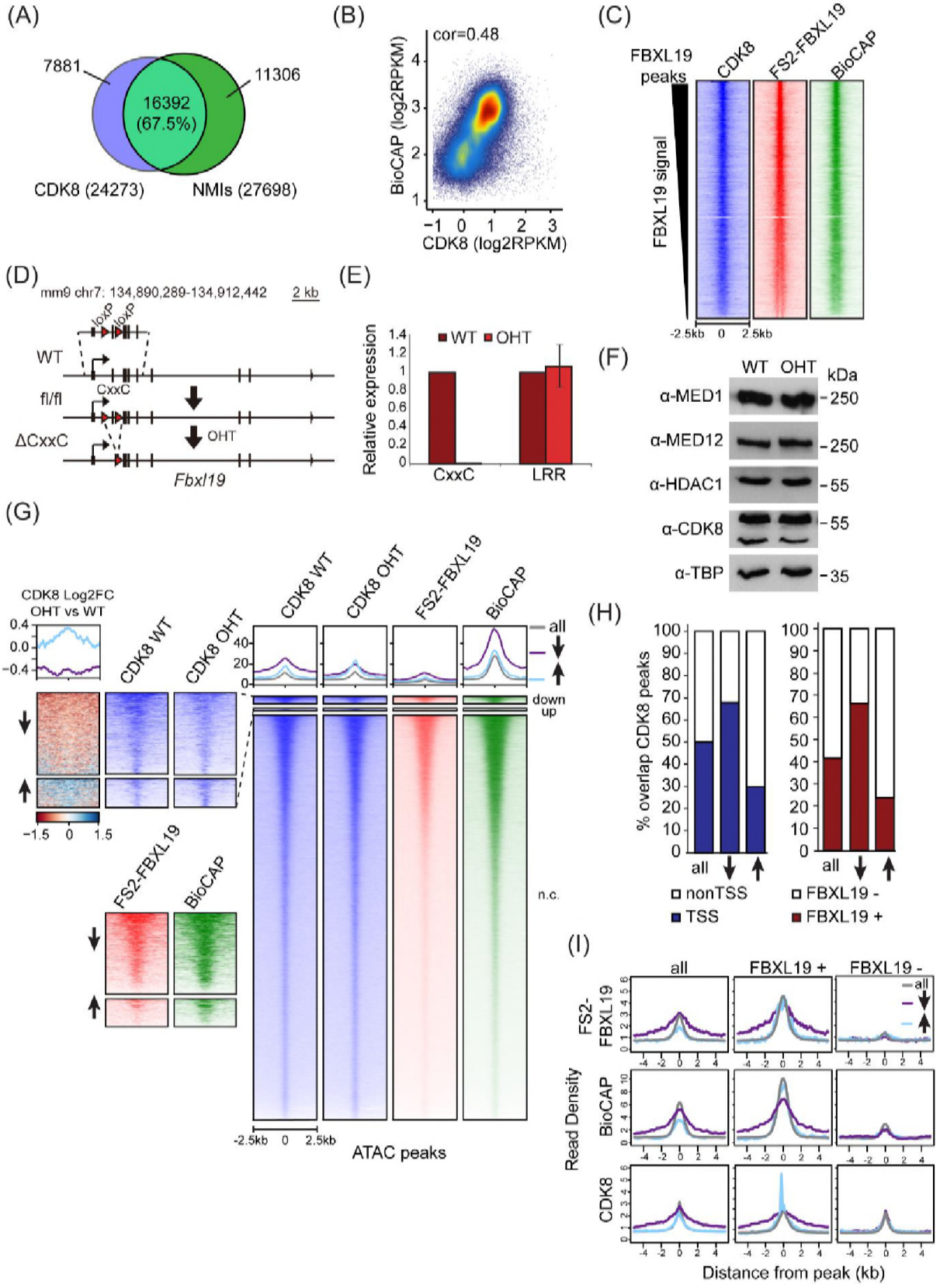
1188

1189 **Figure S3: FBXL19 recruits CDK-Mediator to chromatin.**

1190 (A) Western blot analysis of the expression of the TetR-fusion proteins from nuclear extracts
1191 of the respective cell lines. BRG1 was probed as a loading control.
1192 (B) ChIP-qPCR analysis of the binding of the indicated proteins across the TetO array (distance
1193 from center shown) in TetR-KDM2A (top) and TetR-KDM2B (bottom) ES cell lines. The x-
1194 axis indicates the spatial arrangement of the qPCR amplicons with respect to the center of the
1195 TetO array (in kb). Error bars represent SEM of three biological replicates.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova et al. Supplementary Figure S4.



1196

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure S4: FBXL19 is required for appropriate CDK8 occupancy at a subset of CpG island promoters.

(A) A Venn diagram showing the overlap between CDK8 peaks (n=24273) and NMIs (n=27698).

(B) A scatter plot showing Spearman correlation between CDK8 signal and BioCAP at NMIs.

(C) Heatmaps showing enrichment of CDK8, FS2-FBXL19 and BioCAP at FBXL19 peaks (n=11322) sorted by decreasing FBXL19 signal.

(D) A schematic of the *Fbxl19*$^{fl/fl}$ allele showing the location of the loxP sites. Treatment with tamoxifen (OHT) results in Cre-mediated deletion of the second exon, encoding for the ZF-CxxC domain but leaving the rest of the gene and protein intact.

(E) RT-qPCR of FBXL19 expression in *Fbxl19*$^{fl/fl}$ ES cells before (WT) and following tamoxifen treatment (OHT). Primers specific for the ZF-CxxC and LRR domains were used. Expression is relative to expression in WT ES cells. Error bars show SEM of three biological experiments.

(F) A Western blot analysis of the expression of Mediator subunits in *Fbxl19*$^{fl/fl}$ (WT) and *Fbxl19*$^{\Delta CXXC}$ (OHT) ES cells. HDAC1 and TBP were probed as loading controls.

(G) Heatmaps showing enrichment of CDK8 in *Fbxl19*$^{fl/fl}$ (WT) and *Fbxl19*$^{\Delta CXXC}$ (OHT), FS2-FBXL19 and BioCAP signal at ATAC peaks divided based on change in CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells as in Figure 4E. The mean of the enrichment for each group is shown above the heatmaps. Left: zoomed heatmaps for ATAC peaks associated with a decrease (⬇) or an increase (⬆) in CDK8 binding and a differential heatmap representing log2 fold change of CDK8 signal between OHT and WT samples at the differential peaks.

(H) Percent overlap between CDK8 peaks (as in Figure 4E) and transcription start sites of genes (TSS, left) or FBXL19 peaks (right).

(I) Metaplots showing enrichment of FS2-FBXL19, BioCAP and CDK8 ChIPseq signal at all CDK8 peaks (left), FBXL19-bound CDK8 peaks (FBXL19+, middle) and nonFBXL19 CDK8 peaks (FBXL19-, right). Peaks were divided based on change in CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells as in Figure 4E.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Dimitrova et al. Supplementary Figure S5.

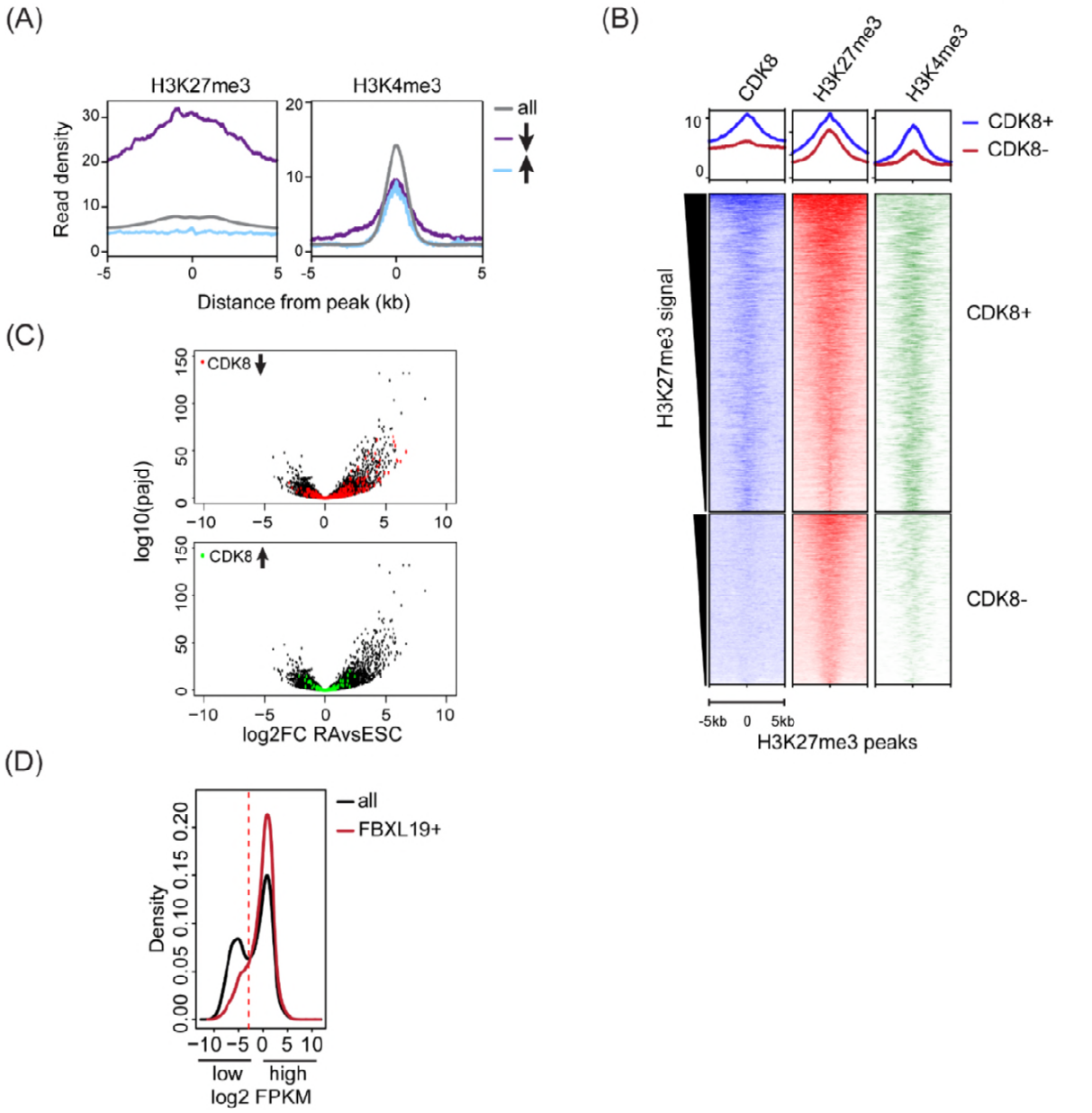


Figure S5: FBXL19 targets CDK8 to promoters of silent developmental genes in ES cells.

(A) Metaplots showing enrichment of H3K27me3 (left) and H3K4me3 (right) ChIPseq signal at all (top), reduced (↓) and increased (↑) CDK8 peaks.

(B) Heatmaps showing enrichment of CDK8, H3K27me3 and H3K4me3 ChIPseq signal at H3K27me3 peaks (n=5588) divided by overlap with CDK8 peaks (CDK8+ n= 3637, CDK8- n=1950) sorted by decreasing H3K27me3 signal. Distance from the peak center is shown.

(C) Volcano plots showing alterations in expression (log2 fold change) comparing ES cells and RA-treated cells. Genes that rely on FBXL19 for normal CDK8 binding in the ES cells state are plotted. Top: reduced levels of CDK8 (red, n=673 of which 203 genes are significantly

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1236 induced by RA and 55 genes are ES-specific). Bottom: increased levels of CDK8 (green, n=255
1237 of which 28 genes are significantly induced by RA and 26 genes are ES-specific).
1238 (D) Genes for Figures 5 D-F were separated into two categories (low and high expression)
1239 based on FPKM value cut-off represented by the dashed line.
1240

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

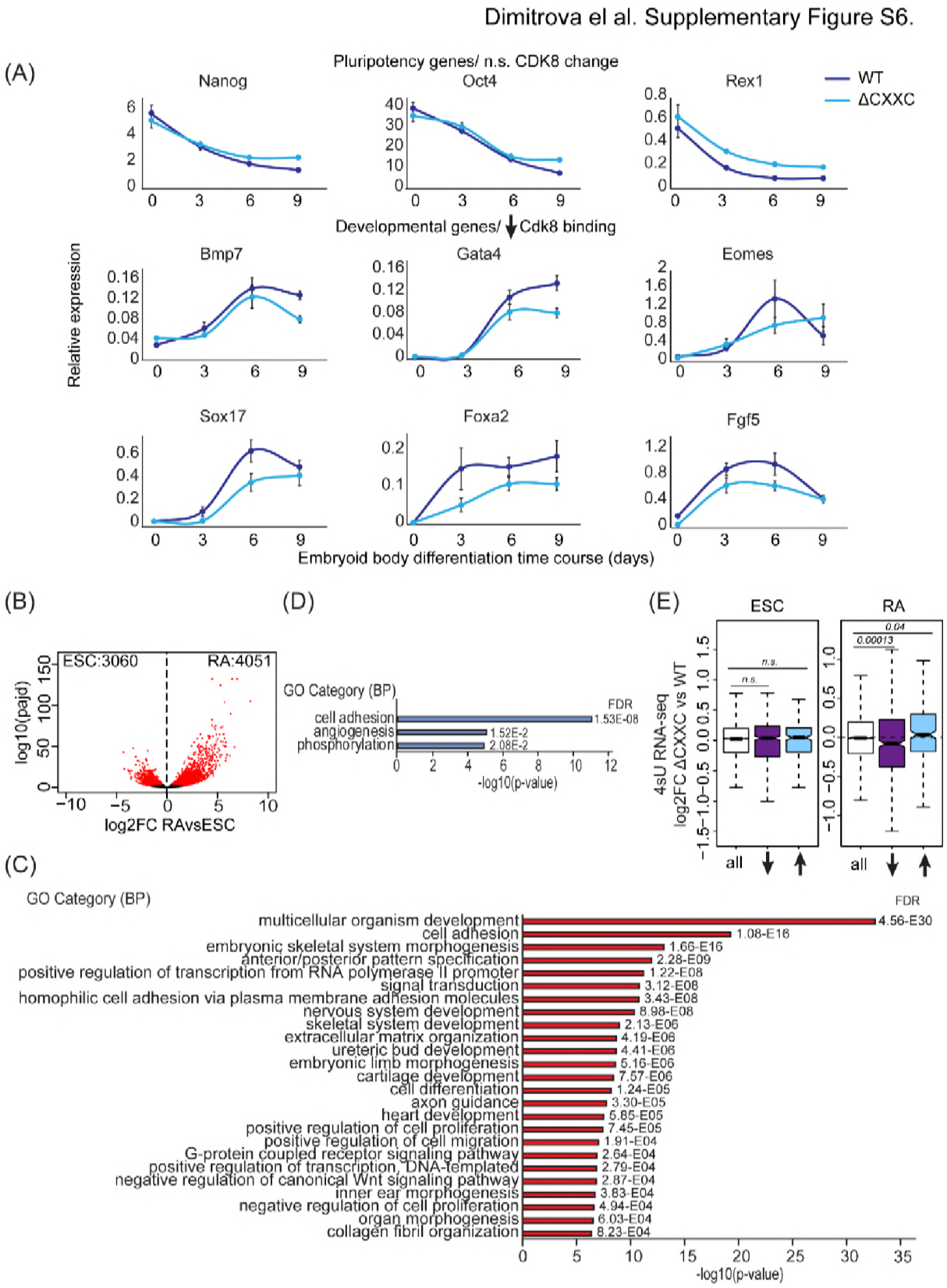


Figure S6: Removing the CpG island-binding domain of FBXL19 results in a failure to induce developmental genes during ES cell differentiation.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(A) RT-qPCR gene expression analysis of developmental genes during an embryoid body differentiation time course. Expression is relative to an average of two house-keeping genes. Error bars represent SEM of three biological replicates.

(B) A volcano plot showing differential expression (log2 fold change) comparing ES cells and RA-induced cells. Differentially expressed genes (log2FC < -0.5 or log2FC > 0.5, padj <0.05) are shown in red. The number of genes considered to be significantly different in expression are indicated.

(C) Gene ontology analysis of the genes induced following RA treatment.

(D) Gene ontology analysis of the up-regulated genes in *Fbxl19*$^{\Delta CXXC}$ cells following RA treatment.

(E) Expression changes of CDK8-target genes (log2 fold change) in *Fbxl19*$^{\Delta CXXC}$ cells compared to WT cells in ES cells (ESC, left) and following RA treatment (right). Genes are divided based on the changes in CDK8 binding in *Fbxl19*$^{\Delta CXXC}$ ES cells - all CDK8 peaks, reduced CDK8 peaks (↓) and increased CDK8 peaks (↑). p-values were calculated using a Wilcoxon rank sum test.

bioRxiv preprint doi: https://doi.org/10.1101/299552; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

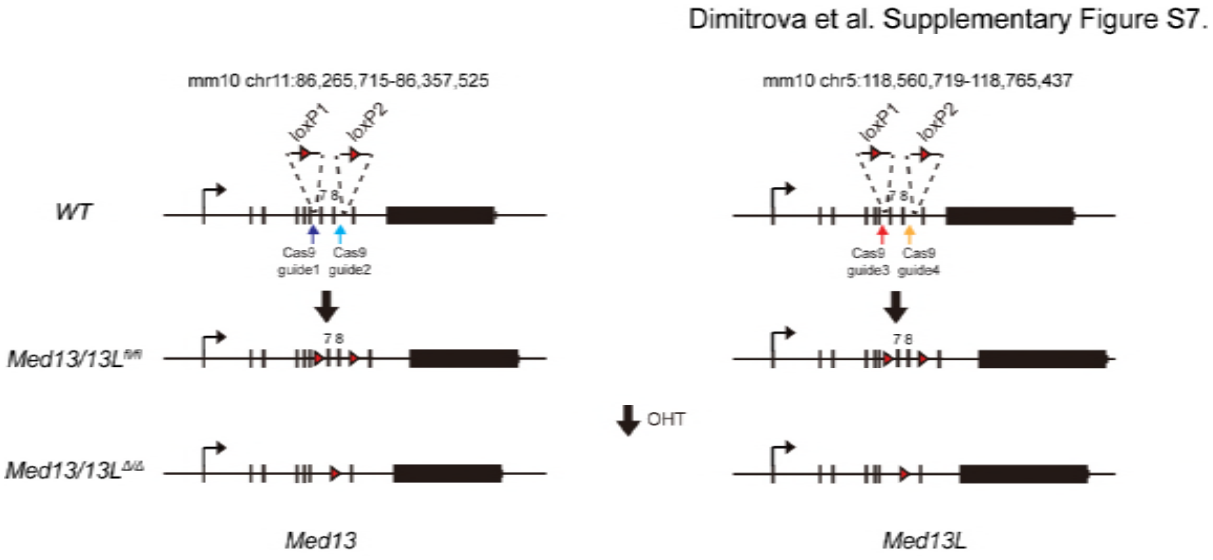


Figure S7: FBXL19 target genes rely on CDK-Mediator for activation during differentiation.

A schematic representation of the strategy used to generate a *MED13/13L* double conditional ERT2-Cre ES cell line. LoxP sites flanking exons 7 and 8 of *Med13* and *Med13L* were inserted using CRISPR-Cas9 mediated targeting. Treatment with tamoxifen (OHT) results in deletion of exons 7 and 8.